



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

The development of a large interval recombinase mediated cassette exchange (RMCE) strategy.

Catherine Penfold

Thesis presented for the degree of
Doctor of Philosophy
University of Edinburgh

2003



Declaration

I declare that the work presented within this thesis is my own,
except where otherwise stated.

A handwritten signature in black ink, reading "Catherine Penfold". The signature is written in a cursive style with a large initial 'C' and a distinct 'P'.

Catherine Penfold

Acknowledgements and thanks

I would like to acknowledge the contributions of various individuals during my study period. Thanks to my supervisor, Andrew Smith, for his advice and suggestions, plus his assistance with some cell culture. Thanks to Doug Higgs and Christina Tufarelli at the Institute for Molecular Medicine in Oxford for their collaboration on this project, their advice concerning the targeting of the alpha-globin locus and generous donations of BAC DNA and the murine alpha-globin annotated sequence data. Thanks to Francis Stewart and his team in Heidelberg for their generous donations of plasmids, specifically Giuseppe Testa for his technical advice. Thanks to Melville Richardson, Helen Wallace and Fabrice Fauer, for their conversations, commiserations and advice on technical matters, specifically with regard to the recET cloning technology. Thanks too, to Karen Johnstone and Stephen Meek for their technical advice. Thanks to Diana Peddie for performing genotyping on my mice and to the members of the animal house staff. Thanks to Derek Rout and Helen Henderson, for maintaining a first class tissue culture facility and to Jan McVicar and the other ISCR members responsible for supporting research, for their efficiency and friendliness. Finally, thanks to the other members of the gene targeting group, past and present, for their good humour and support.

I would like to give my personal thanks to my dear friends and family. Jane, Lubna, Shazia, Anne, Sunnita, Alf and Claire, thankyou so much for your love and support and laughter. Tom, Simon and Marco, thank you for your constructive distractions. Deb, Nan and Kyle thank you for your love and support. Finally, thanks to my parents, whose words have inspired and sustained me.

This thesis is dedicated to Tom, Eddy, Joe and Kate.

Abstract

Murine embryonic stem (ES) cells have provided researchers with a useful tool to investigate genome function and the consequences of genome mutation. One mutational approach is gene-targeting, this involves the introduction of DNA sequences of choice, precisely, to almost any location in the target genome by homologous recombination.

At present, most gene-targeting strategies introduce DNA constructs that derive from plasmids. Plasmids can stably propagate up to approximately 30 kb of DNA. Therefore, this size limit may place a restriction on the range of mutations that may be made to a genome using a single plasmid-derived gene-targeting construct alone. To overcome this limitation, multiple rounds of sequential gene-targeting experiments may be performed, however such an approach may be too lengthy to be practicable. In order to address this current limitation with gene-targeting a novel strategy was tested, implementing Cre-*lox* site-specific recombination (SSR) technology and the bacterial artificial chromosome (BAC) vector system.

Two sequential gene-targeting events in murine E14Tg2a ES cells (HPRT⁻) were performed at separate locations to chromosome 11. The aim of gene-targeting was to create an interval on chromosome 11 that included a single copy of the murine alpha-globin locus, between the hetero-specific *lox* sites, *loxP* and *lox511*, an interval of approximately 64 kb. To this end the first targeting event delivered *lox511/hygromycin/I Sce I/lox511* sequences and the second event *frt/I Sce I/5'hpvt/loxP/neomycin* sequences.

ES cells that were confirmed to have correctly undergone the two desired targeting events (double-targeted) were then assessed to determine whether these events had occurred to the same chromosome 11 (*in cis*), as desired, or to the alternate copies of chromosome 11 (*in trans*). This assessment involved restricting DNA from

the double-targeted ES cell lines with the rare-cutting restriction endonuclease I *Sce* I and resolving the products of this restriction by pulsed field gel electrophoresis. This analysis identified two *in cis* lines (CAT-A3 and CAT-B3) and an *in trans* line (CAT-C10).

The double-targeted ES cell lines were then further characterised to determine whether the hetero-specific *lox* sites they harboured would participate in Cre-mediated SSR. The positive result of this analysis was the generation of ES cell clones that were hemizygous for the alpha-globin locus, a deletion of 64 kb. Hemizygous ES cell clones were obtained from the CAT-A3 and CAT-B3 ES cell lines, as predicted, but not from the CAT-C10 line, although all the lines tested showed evidence of SSR occurring.

In parallel to achieving the interval between *loxP* and *lox511* in ES cells, a BAC, harbouring the alpha-globin locus, was similarly modified with *lox* sites using recombination-mediated cloning. The aim of the BAC modification was to create an interval between *lox* sites in the BAC identical to that achieved in the ES cells. The BAC was targeted sequentially with two separate constructs, *lox511/kanamycin/lox511/HSVtk* and then blasticidin/*loxP/3'hpvt/I Sce I/frt*. The correct targeting of the BAC was verified by restricting its DNA with a panel restriction endonucleases. The *lox* sites were then tested in an *in vitro* analysis with purified Cre recombinase and found to be competent to participate in SSR reactions.

The modified BAC was co-electroporated with a Cre expression plasmid into the CAT-A3 and CAT-B3 ES cell lines, previously characterised as targeted *in cis*, with the aim of exchanging the interval sequences in the ES cell with those of the BAC. The ultimate aim of such an exchange would be to deliver any combination of mutations that would be previously engineered to the BAC interval, to that of the ES cell, by a single SSR event. This experimental approach should expedite and facilitate the mutational analysis of gene loci. To generate comparative data the result of SSR between the modified BAC and an *in trans* targeted ES cell line (CAT-C10) was also assessed.

The selection for the desired exchange involved reconstruction of an *Hprt* minigene and exclusion of a thymidine kinase gene, cells which harboured these events could therefore be selected for in HAT and ganciclovir supplemented media respectively. ES cell clones generated from both of the *in cis* lines tested (CAT-A3 and CAT-B3) had the correct selection resistance profiles, thus indicating that the desired exchange had been achieved in these clones. Additionally, Southern blot analysis from the DNA from these clones was consistent with the achievement of the desired exchange. However, the results obtained from clones generated from the *in trans* line (CAT-C10) were not consistent with their predicted genetic arrangement following SSR with the modified BAC.

Thus far similar experimental approaches have been implemented to exchange smaller intervals of 1 to 5 kb and have been termed recombinase mediated cassette exchange (RMCE). However the experiments described within this thesis are the first test whether the same rationale may be applied to larger intervals. The strategy described and tested in this thesis has therefore been termed large interval RMCE (liRMCE).

Abbreviations

6-TG	6-thioguanine
BAC	bacterial artificial chromosome
bsd	blasticidin
cm	centimetres
DMSO	dimethyl sulphoxide
dH ₂ O	distilled water
ENU.....	ethylnitrosourea
ES (cells).....	embryonic stem (cells)
GANC	ganciclovir
HAT	hypoxanthine, aminopterin, thymidine
hprt	hypoxanthine phosphoribosyl transferase
HR.....	homologous recombination
HSV tk	Herpes Simplex Virus thymidine kinase
hyg	hygromycin
IR.....	illegitimate recombination
IRES	internal ribosome entry site
kan	kanamycin
kb.....	kilobases
kV	kilovolts
msec	milliseconds
μ / m / l	micro / milli / litres
μ g	micrograms
μ F	micro-Farrads
neo	neomycin
oligo	oligonucleotide
PBS.....	phosphate buffered saline
PVP	poly-vinyl-polymer

puro	puromycin
rpm	revolutions per minute
SSR	site-specific recombination

Table of contents

<i>Declaration</i>	<i>i</i>
<i>Acknowledgements and thanks</i>	<i>ii</i>
<i>Abstract</i>	<i>iii</i>
<i>Abbreviations</i>	<i>vi</i>
INTRODUCTION.....	1
1.1 Functional genomics	1
1.2 Genetically modified organisms.....	3
1.2.1 The range of species and uses of GMOs	4
1.2.2 An overview of common methods of genome manipulation used to generate GMOs	5
1.3 Murine embryonic stem cells and gene-targeting	6
1.3.1 Isolating and identifying gene-targeting events.....	8
1.4. The size of gene-targeting vectors	10
1.4.1 The evolution of large vectors and their applications in genome manipulation	11
1.4.2 YAC transgenes and transgenic rescue studies	12
1.4.3 The development of the BAC vector.....	13
1.4.4 BAC transgenesis	15
1.4.5 BAC stability and handling	17
1.4.6 BAC modification by homologous recombination.....	18
1.5 The use of site-specific recombination systems for genome manipulation.	23
1.5.1 Chromosome engineering.....	24
Fig. 1.1 The sequence of <i>loxP</i> and <i>lox511</i> and their diagrammatic representation within this thesis.....	25
Fig. 1.2 Some of the genetic manipulations possible using SSR systems	27
1.5.2 Conditional alleles.....	28
1.5.3 Recombinase Mediated Cassette Exchange (RMCE).....	30
1.5.4 Large interval RMCE	33
1.6 The alpha-globin locus as a test locus	34
Fig. 1.3 The strategy for large interval RMCE in outline.....	35
1.6.1 Diseases associated with the alpha-globin locus	37
Fig. 1.4 Arrangement of the murine alpha-globin locus and genes 5' of the locus	38
1.6.2 Regulation of the alpha-globin locus.....	39
1.6.3 The application of large interval RMCE to study the alpha-globin locus	42
1.7 Thesis outline	45
EMBRYONIC STEM CELL TARGETING	46
2.1 Introduction	46
2.2 The positions of the target sites.....	48
2.2.1 Positioning the target site downstream of the alpha-globin locus	48
Fig. 2.1 Arrangement and nomenclature used for the alpha-globin locus and ES cell recombination cassettes	49

2.2.2 Positioning of the target site upstream of the alpha-globin locus.....	50
2.3 Generating the homology arms for targeting by PCR.....	50
2.4 Targeting the lox511/hygromycin/I Sce I/lox511 construct downstream of the alpha-globin locus	52
Fig. 2.2 Construction of the ES cell targeting vectors CP#47 and CP#48, which contain lox511/hygromycin/I Sce I/lox511	53
Fig. 2.3 Southern blot screening strategy employed to detect targeting of lox511/hygromycin/I Sce I/lox511	54
Fig. 2.4 A southern blot screen of BamHI restricted genomic DNA prepared from a 96 well plate of hygromycin resistant ES cell colonies that arose from electroporation of CP#48 into the E14Tg2a ES cell line.....	55
2.5 Targeting the frt/I Sce I/5'hpvt/loxP/neomycin construct upstream of the alpha- globin locus	56
Fig. 2.5 Southern blotting on putatively targeted clones obtained from electroporation of E14Tg2a with CP#48 (from 96 well plate HYG-4/Plate 2).....	57
Fig. 2.6 Construction of the ES cell targeting vector CP#11, which contains frt/I Sce I/5'hpvt/neomycin	58
2.6 Summary of ES cell targeting	60
Fig. 2.7 Southern blot screening strategy employed to detect targeting of frt/I Sce I/5'hpvt/loxP/neomycin.....	61
Fig. 2.8a A southern blot screen of HindIII restricted genomic DNA prepared from a 96 well plate of neomycin resistant ES cell colonies that arose from electroporation of CP#11 into the HYG-4/Plate2/clone A9 targeted ES cell line	62
Fig. 2.8b A southern blot screen of SacI restricted genomic DNA prepared from a 96 well plate of neomycin resistant ES cell colonies that arose from electroporation of CP#11 into the HYG- 4/Plate2/clone A9 targeted ES cell line	63
Fig. 2.9 Southern blotting on selected double targeted CAT ES cell lines.....	64
CHARACTERISATION OF THE DOUBLE-TARGETED EMBRYONIC STEM CELL LINES.....	65
3.1 Introduction	65
3.2 The strategy used to identify in cis double-targeted ES cell lines	66
3.2.1 Southern blot analysis of restricted CAT genomic DNA separated by PFGE.....	67
3.2.2 Interpretation of the PFGE results	68
Fig. 3.1 PFGE screen to determine in cis or in trans targeting of the targeting constructs	69
Fig. 3.2 PFGE blot hybridised with DnLP	70
Fig. 3.3 PFGE blot hybridised with the neomycin probe	71
3.3 Testing whether the hetero-specific lox sites in the CAT lines are functional....	74
Fig. 3.4 Construction of CP#98, a plasmid used to test the functionality of the lox sites in the double-targeted CAT ES cell lines	75
Fig. 3.5 Functional testing of the lox sites in the recombination cassettes by SSR between CP#98 and an in cis targeted clone.....	76
Fig. 3.6 Functional testing of the lox sites in the recombination cassettes by SSR between CP#98 and an in trans targeted clone.....	77
3.3.1 Experimental details	78
Fig. 3.7 Construction of the Cre expression plasmid, CP#99.....	79

3.4 Interpretation of the results from testing the lox site functionality	80
Fig. 3.8 Summary of the results from electroporation of selected CAT lines with CP#98 and CP#99.....	81
3.4.1 Comparing the efficiency of generating HAT ^R clones between <i>in cis</i> targeted and <i>in trans</i> targeted lines	82
3.4.2 Drug resistance phenotypes from HAT ^R clones generated from the <i>in cis</i> and <i>in trans</i> targeted lines	84
3.4.3 Interpretation of the Southern blot data from the HAT ^R clone generated from the CAT lines assigned as targeted <i>in cis</i>	86
Fig. 3.9 Southern blot analysis from a selected number of HAT ^R clones generated from the co-electroporation of CP#98 and CP#99 into CAT-A3 and CAT-B3 ES cell lines (Experiment III)	87
3.4.3 Interpretation of the Southern blot data from the HAT ^R clone generated from the CAT lines assigned as targeted <i>in trans</i>	88
Fig. 3.10 Southern blot analysis from the HAT ^R clone generated from the co-electroporation of CP#98 and CP#99 into the CAT-C10 ES cell line (Experiment III)	89
3.5 Summary	91
BAC TARGETING.....	93
4.1 Introduction	93
4.2 Design and construction of the BAC downstream targeting construct	95
Fig 4.1 Construction of the <i>lox511/kan/lox511/HSVtk</i> recombination cassette	96
4.3 Targeting the <i>lox511/kan/lox511/HSVtk</i> recombination cassette into the BAC 97	
4.3.1 PCR strategies for preparation of the targeting fragment.....	97
Fig 4.2 Construction of CP#38, the PCR template plasmid used to generate a linear targeting fragment of the <i>lox511/kan/lox511/HSVtk</i> recombination cassette.....	98
Fig. 4.3 The position of PCR primers used to amplify the <i>lox511/kan/lox511/HSVtk</i> recombination cassette.....	99
4.3.2 Restriction digest strategies for the preparation of the targeting fragment.....	100
4.4 Identifying targeted clones within a mixed population of kanamycin resistant colonies arising from the electroporation of the <i>lox511/kan/lox511/HSVtk</i> cassette into <i>E. coli</i> cells harbouring pBAC14567	101
Fig 4.4 Construction of CP#139, the plasmid used to generate an <i>AscI</i> restriction fragment to target the <i>lox511/kan/lox511/HSVtk</i> cassette	102
Fig. 4.5 The restriction enzyme analysis used to determine and confirm <i>lox511/kan/lox511/HSVtk</i> targeting	104
4.5 Checking the integrity of pBAC14567/TK by restriction analysis.....	105
Fig 4.6 Restriction analysis of pBAC14567 targeted with the <i>lox511/kan/lox511/HSVtk</i> recombination cassette to give pBAC14567/TK	106
4.6 Design and construction of the BAC upstream targeting construct	107
Fig. 4.7 Construction of CP#113, the <i>bsd/loxP/3'hppt/I Sce I/frt</i> recombination cassette and BAC targeting vector to target the upstream position of the alpha-globin locus.....	108
4.7 Targeting the <i>bsd/loxP/3'hppt/I Sce I/frt</i> recombination cassette into pBAC14567/TK	109
4.7.1 Screening colonies arising from electroporation of the <i>bsd/loxP/3'hppt/I Sce I/frt</i> recombination cassette.....	109

4.7.2 Elimination of the recombination plasmid by culture with NiCl ₂	110
Fig. 4.8 Restriction digest screening strategy employed to determine <i>bsd/loxP/3'hpvt/I Sce I/fri</i> targeting.....	111
4.7.3 Checking the integrity of pBAC14567/TK/3'HPRT by restriction analysis	112
4.8 In vitro Cre-mediated SSR analysis to confirm the functionality of the lox sites targeted to the BAC.....	112
Fig. 4.9 Restriction analysis of pBAC14567/TK targeted with the <i>bsd/loxP/3'hpvt/I Sce I/fri</i> recombination cassette.....	113
4.8.1 <i>Xho</i> I restriction analysis on <i>in vitro</i> Cre reactions.....	114
4.8.2 Analysis of colonies arising from electroporation of the DNA products of <i>in vitro</i> Cre-mediated SSR	114
4.9 Summary	114
Fig. 4.10 The analysis performed to test the functionality of the <i>lox</i> sites targeted to pBAC14567.....	115
Fig. 4.11 <i>Xho</i> I restriction analysis of pBAC14567, pBAC14567/TK and pBAC14567/TK/3'HPRT following <i>in vitro</i> Cre-mediated SSR	116
Fig. 4.12 The results of replica plating chloramphenicol resistant colonies arising from <i>in vitro</i> Cre-mediated SSR on pBAC14567/TK/3'HPRT	117
LARGE INTERVAL RECOMBINASE MEDIATED CASSETTE EXCHANGE ...	121
5.1 Introduction	121
Fig. 5.1 Replacement of the ES cell locus with the locus residing in the BAC: large interval RMCE.....	122
Fig. 5.2 SSR events to generate either a RMCE event or a co-integrant structure	124
5.2 Experimental Outline	126
Fig. 5.3 A table of the number of HAT resistant clones from each transfection	128
Fig 5.4a Resistance phenotypes of clones derived from the CAT-B3 line	129
Fig 5.4b Resistance phenotypes of clones derived from the CAT-C10 line.....	129
5.3 Analysis of the efficiency with which HAT ^R clones were generated from the CAT-B3 and CAT-C10 lines.....	130
5.4 Analysis on the drug resistance phenotypes of clones generated from the CAT-B3 and CAT-C10 lines.....	131
Fig 5.5 Predicted restriction fragments and probe locations around <i>loxP</i> for <i>in cis</i> and <i>in trans</i> HAT ^R RMCE generated clones	133
Fig. 5.6 Southern blot analysis on clones generated from the CAT-B3 line, with UpRP.....	134
Fig. 5.7 Southern blot analysis on clones generated from the CAT-C10 line, with UpRP.....	135
Fig. 5.8 Checking the integrity of the locus upstream of the RMCE interval for selected clones generated from the CAT-B3 and CAT-C10 lines, with UpLP	136
5.5 Southern blot analysis of clones derived from electroporation of CAT-B3 and CAT-C10 with pBAC14567/TK/3'HPRT	137
5.5.1 The initial screen for the desired recombinants.....	137
5.5.2 Checking the integrity of the genome outwith the RMCE interval	139
Fig 5.9 Predicted restriction fragments recognised by DnLP and DnRP for <i>in cis</i> and <i>in trans</i> HAT ^R RMCE generated clones	140
Fig. 5.10 Checking the integrity of the locus downstream of the RMCE interval for selected clones generated from the CAT-B3 and CAT-C10 clones with DnRP	141
5.5.3 Reconstruction of the <i>hpvt</i> minigene	142

5.5.4 Southern blot analysis to establish evidence of co-integrant structures	142
Fig. 5.11 Southern blot analysis on selected clones generated from the CAT-B3 and CAT-C10 lines, with 3'hpvt probe, neomycin probe and DnLP.....	143
5.6 Summary	145
DISCUSSION	147
6.1 Introduction	147
6.2 Characterisation of the exchanged interval	148
6.3 The effects of undesirable recombination on large interval RMCE.....	149
6.3.1 Genetic arrangements arising from IR followed by SSR	150
Fig. 6.1 Proposed model to generate HAT ^R clones from <i>in cis</i> targeted ES cell lines by IR followed by SSR.....	151
6.3.2 Genetic arrangements arising from a combination of SSR and HR	152
6.3.3 Conditions under which the BAC could participate in HR	153
Fig. 6.2. A diagram to illustrate the generation of HAT ^R clones by the combined events of SSR and homologous recombination with <i>in cis</i> targeted ES cell lines	154
6.3.4 The effects of BAC linearisation due to experimental procedure.....	155
6.3.5 The consequences of a combination of SSR followed by HR	156
MODIFICATIONS TO THE STRATEGY AND APPLICATIONS	160
7.1 Modifications to the large interval RMCE strategy in order to expediate the experimental process	160
7.2 Modifications to the large interval RMCE strategy to study gene loci that can not be accommodated by a single BAC molecule.....	160
Fig. 7.1 A strategy to rapidly produce a variety of mutated ES cell lines	161
7.3 Delivering the hetero-specific lox sites required for large interval RMCE with a single targeting vector	162
Fig. 7.2 Creating a floxed ES cell interval by a single gene-targeting event.....	163
Fig. 7.3 Generating homozygous and hemizygous loci following RMCE	164
7.4 Generating hetero-, homo- and hemizygous mutated loci following large interval RMCE.....	164
7.4 Generating hetero-, homo- and hemizygous mutated loci following large interval RMCE.....	165
7.5 Mediating large interval RMCE in somatic cells	166
7.6 Applying large interval RMCE to functional genomics.....	167
Fig. 7.4 Potential <i>in vitro</i> and <i>in vivo</i> manifestation of RMCE technology	168
7.7 Applying large interval RMCE to orthologous sequence exchange	169
7.8 Applying large interval RMCE to optimise transgene expression	170
7.9 Applying large interval RMCE to investigate <i>in cis</i> regulatory sequences.....	171
7.10 Development of the large interval RMCE strategy in order to recover the exchanged interval.....	173

SOFTWARE, MATERIALS AND METHODS	174
I.1 Software	174
I.2 Chemicals.....	174
I.3 Oligonucleotides	174
I.4 Embryonic stem cells, culture media and solutions.....	174
I.4.1 E14Tg2a ES cells.....	174
I.4.2 Cell culture medium.....	175
I.4.3 Differentiation inhibitory activity (DIA).....	175
I.4.4 Foetal calf serum.....	175
I.4.5 Phosphate Buffered Saline	175
I.4.6 Trypsin solution	176
I.4.7 1 % Gelatin	176
I.4.8 0.1 M 2-mercaptoethanol	176
I.4.9 Freezing Solution	176
I.4.10 Quench Medium.....	176
I.5 ES cell culture	176
I.5.1 Passage and expansion of ES cells.....	177
I.5.2 96 well ES cell passage	177
I.5.3 Freezing ES cells.....	178
I.5.4 Freezing from a small flask.....	178
I.5.5 Freezing a 96-well plate.....	178
I.5.6 Thawing vials of ES cells.....	178
I.5.7 96 well ES cell thawing	179
I.5.8 Electroporation of ES cells	179
I.6 Molecular biology methods	183
I.6.1 DNA Isolation and Purification	183
I.6.2 Enzyme reactions	190
I.6.3 Annealing oligos	193
I.6.4 Quantification	194
I.6.5 Automated sequencing.....	194
I.6.6 Pulsed Field Gel Electrophoresis	195
I.6.7 Nucleic Acid Hybridisation and Screening.....	197
I.6.8 Recombination mediated cloning.....	199
OLIGONUCLEOTIDE SEQUENCES	204
PLASMIDS AND BACS	206
PCR CONDITIONS.....	208
BIBLIOGRAPHY	217

Chapter 1

INTRODUCTION

1.1 Functional genomics

In 2001 the completion of the human genome draft sequences were announced (Venter et al., 2001; Lander et al., 2001). Current estimates on the number of genes range between 30,000 to 45,000 and the functional characterization of each gene is predicted to increase our understanding of fundamental biological processes and aid the development of novel diagnostic and therapeutic methods in medicine. The characterization of gene function in this context is termed 'functional genomics'.

Functional genomics investigations seek to generate data that establish a causative link between specific DNA sequences and the role they play in generating a phenotype. Historically, most functional genomics investigations have relied on the collation of data generated from *in vitro* and *in vivo* studies. However, as whole genome sequencing projects progress and the field of bioinformatics develops more robust tools for sequence analysis, contemporary functional genomics investigations increasingly integrate data from '*in silico*' research. In light of this novel line of investigation the necessity to perform *in vitro* and *in vivo* studies is discussed.

The development of *in silico* approaches now means that it is possible to search the sequence of an entire genome to find sequences that may represent novel genes. Sequence annotation programmes, such as GenomeScan (Yeh et al., 2001),

Genscan (Burge and Karlin, 1997), Genie (Reese et al., 2000) and Genotator (Harris, 1997), aim to predict novel genes *ab initio* using search algorithms based on the recognition of known gene associated sequences. Sequence annotation programmes commonly search for gene associated sequences such as open reading frames and splice-site consensus sequences or aim to identify other specific base compositions that relate to function, such as CpG islands. The collation and integration of data from homology-based studies (e.g. Genewise) and *ab initio* methods (e.g. Genscan) is underway with the Ensembl genome database project which aims to provide a comprehensive annotation of the human genome sequence (Hubbard et al., 2002).

The availability of the entire genome sequence data from a growing number and variety of species, most notably the human and mouse, are continually broadening the capabilities of *in silico* research. At present however, the capabilities of comparative *in silico* functional genomics are unable to distinguish those divergences in sequence homology that are critical to determining phenotype, from those without functional significance. This is especially limiting to the investigation of sequences that control gene expression, such as enhancers and repressors, which remain largely uncharacterised.

The functional consequence of inter-species sequence divergences within regions of coding homology is simplified since there exists an encoded product whose function may be tested. This should determine the effect, if any, that the sequence divergence has on the ability of the encoded product to perform its normal function. In comparison, the ability to evaluate the significance of sequence divergences between non-coding sequences is made more problematic due to the absence of an encoded product. This includes sequences that regulate gene expression and therefore the study of these sequences using an *in silico* approach is less informative.

Furthermore, exclusive of sequences that encode proteins or are known to direct gene expression, the *in silico* approach is unable to identify sequences that may have novel function. This is especially relevant to the discovery and

investigation of sequences that regulate gene expression. Although some enhancers and repressors have been characterised, it is probable that most have yet to be discovered (Lomvardas and Thanos, 2002; Tautz, 2000; Macfarlane, 2000). Therefore without this information it remains impossible to identify these sequences *ab initio*, or perform homology based searches for them. Since a comprehensive knowledge of the elements that affect gene expression is necessary for the complete characterization of genome function, the production of *in vitro* and *in vivo* data are particularly relevant to the discovery and functional characterisation of sequences that control gene expression.

The current *in silico* approaches therefore incur two main shortcomings: a reliance on existing data to characterise the function of a sequence and a dependency on the hypothesis that homology between sequences indicates shared function. The computational assignment of gene sequences is therefore currently regarded as putative. The assignment of function to a sequence that has been identified by *in silico* methods is still most reliably validated by the generation of data from conventional *in vitro* and *in vivo* studies. Therefore the genetically modified organism will remain as a key tool for investigative functional genomics and validation in the emerging *in silico* research era.

1.2 Genetically modified organisms

The most accurate approach to the investigation of genome function is to observe how the genome determines phenotype within its natural *in vivo* environment. However, in some cases, for example the investigation of the human genome, this approach may not always be ethical or practicable. This becomes especially relevant if experimentation is required to verify a hypothesis, or if the system under study is so complex that it require subdivision to clarify the analyses. To overcome these obstacles model systems, of whole organisms or cell lines, are used. The implementation of a model system often permits the investigator to evaluate genome function within a simpler context and in a system that is more amenable to experimentation.

An organism with an altered genome is termed a genetically modified organism (GMO). In terms of technology, there are two main features of a GMO. The technology implemented to manipulate the genome (as discussed later in Section 1.2.2) and the technology applied to generate an *in vivo* phenotype from the altered genome (see below).

1.2.1 The range of species and uses of GMOs

GMOs have been generated from an extensive range of prokaryotic and eukaryotic species, at both the uni- and multi-cellular level. Species that are popular to model human genetics usually share equitable physiology, however the application of GMOs is not limited to this area. The application of current GMO species include: creating models to test pharmacologicals (Rudolph and Mohler, 1999); xenotransplantation (Logan, 2000); creating bioreactors (Larrick and Thomas, 2001; Janne et al., 1994); the creation of modified animal and plant species for consumption and nutrition (van der Meer et al., 2001); in addition to creating models to study gene function and human disease (Jonkers and Berns, 2002; Bucan and Abel, 2002; www.mgu.har.mrc.ac.uk; www.biomednet.com/databases/currbiol/mko/dataset.exe). In addition to plant species other species that are commonly utilised for some of the purposes listed above include: *Bos Taurus* (Piedrahita, 2000); *Brachydanio rerio* (Patton and Zon, 2001); *Caenorhabditis elegans* (Jorgensen and Mango, 2002); *Drosophila melanogaster* (St Johnston, 2002); *Mus musculus*; *Ovis aries*; *Rattus rattus* (Jacob and Kwitek, 2002); *Sus scrofa* (Wheeler and Walters, 2001); *Tympanuchus cupido* (Sang, 1994); *Xenopus laevis*.

The mouse (*Mus musculus*) has, in particular, proven to be an attractive *in vivo* model. In addition to its equitable physiology and ease of animal husbandry, the other appealing feature of the mouse as an animal model is its genetics. The mouse has a high sequence identity and orthologous sequence alignments to the human genome. Recently the draft sequence of the mouse genome was published and this powerful new resource marks a new era in the way in which experimental genetics can be done.

The generation of a murine line or murine cell line harbouring an altered genome may be performed by a variety of methods, including: microinjection of a single-cell embryo (Gordon and Ruddle, 1981); embryonic stem (ES) cell contribution within a blastocyst (Gossler et al., 1986; Robertson et al., 1986). The research efforts that focussed on transferring these two methods to other species were largely unsuccessful. However, with the emergence of nuclear transfer technology (Campbell et al., 1996) it is now theoretically possible to extend the production of GMOs to any species that produces oocytes, thereby broadening the application of GMOs for medicine and research (Colman, 1999).

1.2.2 An overview of common methods of genome manipulation used to generate GMOs

The other necessary component for the production of a GMO is genome manipulation. Methods of genome manipulation are extremely varied but can be discussed in terms of the level of control the investigator retains on determining the mutation generated. Mutagenic approaches, that are completely random, exert control over neither the manipulation made, nor its location. Examples of random mutagenic approaches include the use of radiation, such as X-rays (Steward and Nusslein-Volhard, 1986), or chemical mutagens, such as ENU (Haffter et al., 1996; Nolan et al., 2000). Less random approaches control the manipulation that is delivered to the genome but not the location of the manipulation, for example the use of viral (Boris-Lawrie and Temin, 1993; Kovesdi et al., 1997), retroviral (Hanazono et al., 2001; Janson and Doring, 2001) and transposon (Izsvak et al., 2000) based vectors and transgenes. Less random still is the use of gene-trap vectors (Cecconi and Meyer, 2000), whose initial application was the identification and report of transcriptionally active genes and so control over the manipulation and some control over the location is retained. Ultimately though, gene targeting provides the ability to control both the location and the nature of the manipulation made. Gene targeting describes technology to direct sequences of choice, precisely, to almost any location in the genome, in cells that support homologous recombination, for example murine embryonic stem cells.

Therefore ES cells are particularly useful, their ability to undergo homologous recombination and support gene targeting makes their genome amenable to genetic modification. Subsequently, generating a mouse model organism can then be used to assess the phenotypic consequences of modifications to the genome.

1.3 Murine embryonic stem cells and gene-targeting

Murine embryonic stem (ES) cell lines are immortal and can be derived from the inner cell mass of pre-implantation blastocysts (Evans and Kaufman, 1981; Martin, 1981). ES cells are termed pluripotent due to their unique ability to become any cell type of the developing embryo and eventually adult organism (Nagy et al., 1990). The ability of ES cells to contribute towards the germ-line in chimaeric mice was first demonstrated in 1984 (Bradley et al., 1984). The subsequent inheritance of the ES cell genome can be achieved by breeding germline chimeras, first to generate heterozygous mice and then by inter-cross to generate homozygous animals.

In addition protocols have been developed to differentiate ES cells along a variety of cell lineages (for reviews see: Rathjen and Rathjen, 2001; Wobus, 2001; Rossant, 2001). For example, neuronal (Trophepe et al., 2001), cardiomyogenic (Kolossoff et al., 1998), hepatic (Jones et al., 2002) and hematopoietic cell lineages (Suzuki and Nakano, 2001) may be derived. Alternatively somatic cell lineages may be derived from transgenic mice generated from ES cells. The first report of using ES cells to transmit a manipulated genome through the germline was in 1986 (Gossler et al., 1986) and this rationale now forms the bases for the routine generation of a large proportion of murine GMOs. Thus, the investigation of the manipulated ES cell genome can be diversely studied, both *in vivo* and *in vitro*.

As outlined previously, many strategies of genome manipulation exist, with each providing the investigator various levels of control over the manipulation made (see Section 1.2.2) and most of these strategies may be applied to manipulate the ES cell genome. However, it is the ability of murine ES cells to mediate

homologous recombination that makes them amenable to the most precise form of genetic manipulation currently possible, namely gene targeting.

The gene-targeting process is commenced with the design and production of a DNA molecule capable of integrating to a specific location within a genome. This DNA is commonly termed a 'targeting vector' or 'targeting construct' and the location to which it is targeted may be known as either the 'target' or 'target site'. The targeting vector integrates specifically to its target in the genome by virtue of regions of sequence homology contained within the targeting vector to the target site. These regions are often termed 'homology arms' and enable the targeting vector to participate in homologous recombination. The integration of a targeting vector to a target site by homologous recombination is known as 'targeting'.

Gene targeting in ES cells was first demonstrated in 1987 to create a functional HPRT gene (Thomas and Capecchi, 1987; Doetschman et al., 1987) and again in 1988 to mutate HPRT (Doetschman et al., 1988), a manipulation that was subsequently the first demonstration of a germline transmission of a gene-targeted event (Koller et al., 1989).

It is possible to precisely integrate targeting constructs to most regions of the genome, although the efficiency with which targeting constructs integrate may vary. The main factor that reduces gene-targeting efficiency is the propensity of the targeting vector to undergo random integration into the target genome, by a process of illegitimate recombination. The integration of the targeting vector to any other genome location other than its target can be considered as a random integration event.

The likelihood that a targeting vector will integrate at random is largely unpredictable but may be affected by its composition. The most influential factor is the nature and extent of the homology that exists between the targeting vector and the target site. It has been shown that any decrease of isogenicity between the homology arms and target site sequences can reduce targeting efficiency in ES cells (te Riele et al., 1992). The length of the homology arms also has an effect, with larger arms usually increasing targeting efficiency (Deng C, 1992).

Therefore, following the introduction of the targeting vector to the ES cells, usually by electroporation, it is necessary to determine whether targeting has occurred or whether the targeting vector has integrated to a random genomic location, or indeed whether the genome remains unchanged.

1.3.1 Isolating and identifying gene-targeting events

There are several possible outcomes of introducing a gene-targeting construct to ES cells by electroporation. The desired outcome is of course the targeted integration of the gene-targeting construct to the required target location, mediated by homologous recombination. However, alternatives to this outcome include cell death, or the generation of untargeted cells, by either the gene-targeting construct not entering the cell or the inability of the gene-targeting vector to stably integrate into the genome once inside the cell. Alternatively, the gene-targeting vector may integrate into the cell genome, not at the target site as desired, but instead at a random location by illegitimate recombination. There are various ways to ensure that cells harbouring these undesired events are eliminated in order to isolate a population of ES cell clones that harbour the desired targeting event.

In order to distinguish those cells that have integrated the gene-targeting vector into their genome from those whose genome remains unchanged, it is possible to include a positive selection marker in the vector. These markers usually take the form of a gene and are included in the gene-targeting vector during its construction. The positive selection marker is usually located between the homology arms of the gene-targeting vector, in order to maximise the probability of its incorporation to the genome during targeting. The incorporation of the gene-targeting vector into the target genome permits the stable expression of the positive selection marker gene and this confers resistance to a supplement (usually an antibiotic) that is subsequently added to the cell culture media following electroporation. Therefore those cells that harbour the gene-targeting vector in their genome survive the addition of the supplement and those cells without the vector die.

The gene-targeting vector may also integrate undesirably to a random location in the target genome by illegitimate recombination. Illegitimate recombination events may integrate the entire gene-targeting vector and in these cases the use of a positive selection marker does not distinguish those cells that have targeted from those cells that harbour a random integration of the gene-targeting vector. The analysis of DNA derived from clones in such a mixed population may determine those that are targeted from those that are not. Analyses such as PCR around the target site or a Southern blotting screening strategy would indicate whether the gene-targeting construct was present at the target site or not, since the DNA from clones that harbour a random integration event would appear unchanged at the target site.

However the ability to target some locations in the genome may be less feasible than for others. Genome locations that are less amenable to gene-targeting may therefore be difficult to identify with a positive selection strategy alone, due to an overwhelming number clones generated by random integration events. To overcome this problem it may be possible to screen a large number of clones (500 to 1000) in order to identify a targeted clone in the population, however a more practical approach is to use a strategy of positive negative selection (PNS). PNS employs a positive selection marker (as described above) and another marker, usually a gene, whose presence within the genome can be selected against. In other words, a supplement to the cell culture media is able to kill those cells that harbour the negative selection marker.

The successful outcome of a targeted integration of a PNS gene-targeting vector is the integration of the positive selection marker into the genome and the exclusion of the negative selection marker. This is achieved by placing the negative selection marker outwith the homology arms during the construction of the gene-targeting vector. Therefore upon homologous recombination those sequences between the homology arms are integrated into the genome (i.e. the positive selection marker) and those outside the homology arms are excluded from integration (i.e. the negative selection marker). Since random integration by

illegitimate recombination usually integrates the entire gene-targeting vector into the genome then in these instances both the positive and negative selection marker are integrated. Thus, when selection is applied to the cell culture media those cells harbouring the negative selection maker, and therefore a random integration of the gene-targeting vector, die and those cells harbouring a targeted integration survive.

A PNS strategy therefore enriches a population of clones that harbour the gene-targeting vector for those that harbour the vector at the target site, and although enrichment may not be absolute (between 80 and 100 %), the PNS strategy makes it easier to identify the desired targeted clones than with a strategy using positive selection alone. Additionally, PNS may be used as a means of distinguishing the level of homologous recombination from the level of illegitimate recombination for the specified targeting construct.

An example of a typical PNS strategy involves the construction of a gene-targeting vector with the positive selection marker situated between the homology arms of the vector and the thymidine kinase gene as the negative selection marker, outside the arms. Following electroporation, the positive marker is first selected for, which isolates the population of cells that have incorporated the targeting vector into their genome, whether targeted or random. Next, those cells that harbour a random integration of the gene-targeting vector are removed from the population by their sensitivity to the addition of ganciclovir (GANC) to the ES cell media, due to the presence of the thymidine kinase gene in their genome (Mansour et al., 1988). This ability to select against the presence of the thymidine kinase gene has been used in the design of the strategy presented within this thesis.

The other selection strategy relevant to the strategy presented within this thesis is the reconstruction of the HPRT minigene, which is discussed in Section 1.5.1 (a).

1.4. The size of gene-targeting vectors

To date, most gene-targeting constructs have been plasmid derived. This is largely due to the historical, robust and widely practiced methods for the

propagation and manipulation of plasmid vectors. Conventionally DNA vector construction in *E. coli* has been largely mediated with restriction endonuclease based methods (Sambrook et al., 1989) and this remains the case, although recently methods that implement PCR based cloning techniques have emerged (Akiyama et al., 2000).

The use of plasmid vectors restricts the size of DNA that may be propagated in a single molecule. The upper threshold for the stable and efficient replication of most plasmids is approximately 25 to 30 kb. The use of plasmid derived targeting constructs therefore places restrictions on the design of the targeting construct, which in turn restricts the range of manipulations that may be made to the ES cell genome by gene-targeting alone.

1.4.1 The evolution of large vectors and their applications in genome manipulation

Transgenesis describes evolving technology that mediates the production of a transgenic organism. The term ‘transgenic’ was initially used to describe adult mice that had inherited an altered genome in every cell by the germline transmission of additional gene sequences or ‘transgenes’ (Gordon and Ruddle, 1981; Costantini and Lacy, 1981). However, the addition of gene sequences is not a pre-requisite for the term transgenic, since organisms that have had gene sequences deleted or replaced (Detloff et al., 1994) (known as ‘transgenic knock-outs’ (Barinaga, 1994), or re-located (Buchholz et al., 2000) have been described as transgenic.

The early transgenic studies of the 1980s used randomly inserted plasmid based vectors (up to 15 kb in size). These revealed that both the components of a transgene and the nature of the genome surrounding its insertion site could modify the transgene’s expression. The effects that the chromosomal location may have on the expression of a transgene have since been termed ‘chromosomal position effects’ and they present a major obstacle in generating the desired expression from a transgene.

During the 1980s studies were beginning to elucidate genomic sequences that appeared to act *in cis* and could regulate gene expression at a distance. These sequences were termed ‘enhancers’. It was therefore hypothesized that the absence of enhancer sequences from a transgenic vector was the factor responsible for the lack of robust transgene expression at any given locus. To overcome this problem two approaches were taken, either the gene of interest was studied and mutated at its normal position in the genome (with gene-targeting approaches) or vectors larger than plasmids were employed as transgenes.

1.4.2 YAC transgenes and transgenic rescue studies

In an effort to overcome chromosomal position effects with small randomly integrated transgenes, vectors that were capable of maintaining large genomic inserts were tested for their ability to function as transgenes. It was hoped that by using large vectors, regions of sequence, that would normally flank the transgene at its usual genome location, would be integrated along with the transgene and that these sequences would include all of the required enhancer elements necessary for the transgene’s expression. The aim of using such large vectors was to achieve position-independent expression of the transgene. In addition it was anticipated that this approach would facilitate the discovery of distant regulatory sequences, since the transgene provided a discrete region of genomic sequence for analysis.

Early attempts at implementing transgenesis in mice with larger vectors achieved the desired position-independent expression, initially using YACs as the vector of choice (Forget, 1993; Lamb and Gearhart, 1995; Peterson et al., 1997; Peterson et al., 1995). In addition to achieving robust transgene expression, YAC transgenesis also became the method of choice to study enhancers and other elements that affected gene expression at a distance *in cis* (Peterson et al., 1993). The integrated YAC transgene provided a discrete genomic region that could be evaluated for sequences necessary to provide position-independent expression. Concurrently, gene-function was being investigated by producing knock-out and other mouse mutations and investigating the resulting phenotype. The YAC transgene approach was used in conjunction with knock-out and mutation studies to

determine whether sequences in the transgene were capable of reversing (or rescuing) the mutated phenotype to generate a wild-type animal. These strategies were termed ‘transgenic rescue experiments’. The extent to which a transgene was able to rescue a mutated phenotype, was not always complete, giving rise to partially-rescued phenotypes (Lakshmanan et al., 1998; Slee et al., 1999). Therefore the value of a ‘rescue experiment’ is dependent on rigorous experimental analysis in order to verify the process. This includes the characterisation of the mutation that causes the mutated phenotype, which is made more problematic if the mutation is unmapped or generated by the random mutation of the test genome. Following characterisation of the mutation the rescued phenotype requires rigorous assessment, to ensure that all of the mutated characteristics are reversed.

However, the most significant disadvantage related to the use of YACs for transgenic rescue experiments is the amount of characterisation that must be performed in order to determine the nature of the transgene insertion. It is necessary to determine whether any undesirable re-arrangements or fragmentation of the YAC have occurred, and also to determine whether the transgene is present as a single copy, both at the site of integration and genome wide. The reason for undertaking such extensive characterisation is largely due to the inherent instability of the YACs molecule (as discussed later) and can therefore be overcome by using alternative large vector transgenes.

The alternative vectors to the YAC, that are competent to propagate large genomic inserts include: the P1 artificial chromosome (PAC); human or mammalian artificial chromosomes (HACs or MACs); and bacterial artificial chromosomes (BACs). The following section discusses BACs and their advantages over the other artificial chromosome systems, most notably YACs, since, thus far, YACs have been the large vector of choice for genome manipulation.

1.4.3 The development of the BAC vector

The BAC system is based on the *E. coli* F-factor, which is a single copy plasmid that exists in supercoiled circular form in host bacterial cells (Shizuya et al., 1992). There are three main advantages of using BACs rather than other large

vector systems, the first is the availability of genomic libraries from a wide variety of species, secondly the relative stability of BACs makes them easy to handle in the laboratory and applicable for high-throughput systems and thirdly since BACs are harboured in *E. coli* this renders them amenable to sequence modification by homologous recombination, see Section 1.4.5

The development of the BAC system was in part driven by the Human Genome Project (HGP) as a means to easily construct robust and manageable genomic libraries and physical maps for genomic sequencing (Monaco and Larin, 1994). The ability to construct libraries in BAC vectors that ensure coverage of the entire genome in question with individual clones that faithfully represent the original genome sequence was therefore an important advance.

Before the development of BACs, YAC libraries were commonly used for genome mapping. However the large percentage (40 to 50%) of chimaeric (Green et al., 1991; Libert et al., 1993) and unstable (Dunford et al., 1993; Neil et al., 1990; Schmidt et al., 1994) YAC clones hindered the usefulness of these libraries. Although it has since been shown that BAC clones containing tandemly repeated DNA may be highly unstable (Song et al., 2001) and the development of recombination deficient YAC host strains does reduce the levels of chimerism and re-arrangement that occur (Chartier et al., 1992; Haldi et al., 1994; Kohno et al., 1997; Palmieri et al., 1997).

BAC libraries are commonly produced by cloning single fragments generated either by the products of whole genome partial restriction or physical shearing into a replicative BAC vector backbone. The current generation of BAC libraries produce BACs that average in size between 100 kb to 150 kb. However, BACs have the capacity to maintain much larger fragments of genomic DNA, for example, during the characterisation of a chicken BAC library, the largest BAC that was identified stably maintained a cloned insert of 700 kb (Zimmer and Verrinder Gibbins, 1997).

A number of BAC cloning vectors are available. The addition of the *lacZ* gene to the original pBAC108L created pBeloBAC11 and its derivative pECBAC1.

*Eco*RI fragments are easily cloned into pECBAC1 and pBeloBAC11 accepts *Hind*III to *Bam*HI fragments. The second generation of pBeloBAC11 vectors includes pEBAC1 and pEBAC140 and the Rosewell Park Cancer Institute (RPCI) uses BAC vector pBACe3.6. BAC libraries have been constructed from a wide variety of species, with databases accessible from: <http://informa.bio.caltech.edu/>; <http://www.resgen.com/intro/libraries.php3>; <http://hbz.tamu.edu/bacindex.html>; <http://www.chori.org/bacpac/libraryres.htm>.

1.4.4 BAC transgenesis

A number of studies have now obtained functional data from transgenic BAC studies. Most of these studies have been conducted with transgenics produced from the random integration of microinjected BAC DNA into fertilized mouse oocytes. The first reported experiment of this type being analysis of the Clock genes for circadian rhythm (Antoch et al., 1997). Since this study other groups have found success using BAC vectors to rescue mutations.

BAC transgenesis has been implemented in the investigation of tissue and organ development, e.g. the 'Granulation rescue and developmental marking of juxtaglomerular cells using "piggy-BAC" recombination of the mouse ren locus' (Mullins et al, 2000). This study exemplifies the use of BAC transgenes not only to rescue a mutated phenotype but also mark a set of tissues in order to study their progression throughout development. The mutant phenotype for the homozygous Ren-1 null knock-out mouse was successfully rescued by a 145 kb BAC transgene. The BAC contained the Ren-1 gene, which had been previously altered by homologous recombination in *E. coli* to introduce the IRES- β geo sequences, in order to mark those cells that express Ren-1. The implementation of a breeding programme, between Ren-1 null mutants and mice homozygous for the inserted BAC transgene, produced *in vivo* mouse models with a range genotypes. The resulting phenotypes generated a comparative data set that furthered the understanding of the function of the Ren-1 gene.

BAC rescue has been applied to aid the mapping of genes within naturally occurring deletions in mice, e.g. 'Physical localization of the mesoderm development (*mesd*) functional region' (Wines et al., 2000). The investigators in this study took a subset of deletions on mouse chromosome 7 that were known to affect mesoderm development and isolated a number of murine BACs to create a contig library that spanned this region, a distance of approximately 1.3 Mb. The investigators then tested the BACs to see which of them was capable of rescuing the mutant phenotype. This resulted in a discrete region of approximately 75 kb being defined, which was then investigated for expressed sequences. This generated candidate genes responsible for rescuing the mutant phenotype. In this way the identification and mapping of a candidate gene responsible for a mesoderm development, within a 1.3 Mb deletion, was simplified and facilitated by BAC transgenic rescue.

BAC transgenesis has also been implemented to verify candidate genes thought to be responsible for human diseases, e.g. 'TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome' (Merscher et al., 2001). Velo-cardio-facial/DiGeorge syndrome is caused by deletions of 1.5 Mb to 3 Mb in humans at 22q11. The deletion of a candidate gene, *TBX1*, was thought to play a crucial role in generating the human disease phenotype. However, without a naturally occurring deletion in mice that mimicked the human disease phenotype this theory was difficult to test. Therefore the investigators engineered the equivalent of the 22q11 deletion, using *Cre/lox* technology, in mice. They engineered a 1.5 Mb deletion on mouse chromosome 16 at the locus equivalent to the human deletion, this resulted in the production of a number of hemizygous genes. A BAC containing the human sequence of the candidate disease-related gene, *TBX1*, was introduced to these mice by breeding. The BAC partially rescued the disease phenotype in the mice, thereby providing support to the theory that the deletion of *TBX1* contributes to the velo-cardio-facial/DiGeorge disease phenotype in humans.

BAC transgenesis has been implemented to determine the dominance of disease loci, e.g. 'Transgenic rescue of the tattered phenotype by using a BAC encoding Ebp' (Means et al., 2001). The tattered phenotype in mice is thought to be equivalent to the mutation that causes the X-linked dominant disease, chondrodysplasia punctata (CDPX2). Human females that are heterozygous at the CDPX2 locus, display skin, hair, craniofacial and skeletal abnormalities and cataracts, but affected males are not observed. This pattern suggests that the disease is prenatal lethal in the hemizygous state in humans, as is the tattered phenotype for mice. The molecular basis of the phenotype in humans and mice was established as arising from a mutation of the emopamil-binding protein gene (Ebp) and the investigators of this study wished to eliminate the possibility that this mutation was dominant-negative. The investigators generated data that demonstrated that the transgenic expression of a wild-type murine Ebp protein, from a single 125 kb murine BAC transgene, was sufficient to fully rescue the tattered phenotype of males and heterozygous females, which carried the mutated Ebp gene. This observation argued against a dominant-negative mechanism in heterozygous tattered females and verified the role of Ebp in generating the disease phenotype.

The studies detailed above implement transgenesis by introducing the BAC DNA by oocyte micro-injection. Thus far however, there has been no report of transgenesis arising from the electroporation of BACs into ES cells, at either random genomic locations or as part of a gene-targeting strategy.

1.4.5 BAC stability and handling

The stability of BAC DNA is in part due to the fact that they are harboured in recombination-deficient *E. coli* host strains. Most commonly the DH10B strain is used, whose genotype is: F⁻*araD139* Δ (*ara*, *leu*)7697 Δ *lacX74* *galU* *galK* *rpsL* *deoR* ϕ 80*dlacZ* Δ M15 *endA1* *nupG* *recA1* *mcrA* Δ (*mrr* *hsdRMS* *mcrBC*) (Grant et al., 1990). The presence of the *mcrA* genotype marker and deletion of *mcrBC* and *mrr* makes this strain suitable for cloning DNA that contains methylcytosine and methyladenine. DH10B is receptive to large vectors due to the presence of the *deoR*

mutation. Therefore, both prokaryotic and eukaryotic genomic DNA can be cloned effectively into vectors harboured in DH10B.

BAC DNA can be reliably purified from the *E. coli* bacterial chromosome using methods similar to those for plasmid isolation, although yields of BAC DNA are significantly lower due to their single or low copy status within the host bacterial cell. In contrast, it is difficult to isolate YACs free of yeast genomic DNA, giving rise to problems during YAC library production (as mentioned previously) and the introduction of YAC DNA to ES cells (Choi et al., 1993; Davies et al., 1992; Davies et al., 1993; Jakobovits et al., 1993; Lamb et al., 1993; Pearson and Choi, 1993; Strauss et al., 1993) or oocytes (Montoliu et al., 1994; Schedl et al., 1996; Schedl et al., 1993).

Since the average BAC size is less than 200 kb it may be possible to analyse the cloned insert by restriction and gel electrophoresis alone, thereby obviating the requirement for pulsed field gel and Southern Blot analysis, which are a necessity for larger vector constructs such as YACs and HACs. The ease with which BAC DNA can be maintained, purified and analysed contributes a significant advantage of BACs over YACs and HACs. The BAC has therefore become the vector of choice for genome sequencing and whole genome library contig assembly using STS and EST mapping (Hubert et al., 1997; Johnson et al., 1999; Kim et al., 1996; Osoegawa et al., 2000) due to the practicable scale-up of BAC propagation and analysis for high-throughput applications.

1.4.6 BAC modification by homologous recombination

BACs are harboured in the *E. coli* bacterium and so methods of homologous recombination mediated cloning can be implemented to manipulate any part of the vector or cloned insert. Homologous recombination has been routinely used to manipulate YACs (for a review see Duff and Huxley, 1996), therefore the establishment of comparable techniques for BAC modification was necessary to enable their equivalent application. This is currently made possible by recombination-mediated cloning.

In the broadest terms recombination mediated cloning utilizes homologous recombination pathways, in various strategies, to manipulate any DNA molecule that can be propagated within the *E. coli* bacterial cell. These include: the *E. coli* chromosome; plasmids; BACs and PACs. These approaches do not rely on the presence of suitable restriction sites and can therefore be used to insert, delete or substitute DNA sequences at any desired position on the target DNA molecule. Consequently, the limitations that apply to conventional cloning techniques are alleviated. Several methodologies have evolved to perform recombination mediated cloning, however these can be categorized according to the recombination pathway that is used, namely *recA* or *recE/recT* or red.

In *E. coli* the major pathway of general homologous recombination mechanism is initiated by cooperation between RecA (the most conserved strand invasion protein in evolution) and the multi-functional RecBCD enzyme complex, which is capable of vigorous exonuclease activity (Kowalczykowski et al., 1994).

Methods that utilize *recA* recombination are based on two-step modification protocols. The application and efficiency of this approach is dependent on using large regions of homology, 400 bp or greater, harboured on a purpose-built targeting plasmid, the efficiency of the counter selection procedure and the nature of the provision of RecA. The *recA* gene could be omitted if a RecA⁺ *E. coli* host strain is used, although providing RecA *in trans* from the targeting plasmid permits the use of common BAC hosts that are *recA*, such as DH10B, and also limits the recombinogenic window thereby limiting the possibility of any undesirable rearrangements in the target molecule.

X.W. Yang first demonstrated the modification of a BAC by a two-stage *recA* mediated method utilizing a temperature sensitive targeting plasmid and *tet* counter selection (Yang et al., 1997). RecA-dependent BAC modification has since been applied to study the murine renin locus (Payne et al., 1999; Mullins LJ, 2000) and improved by using *rpsL* counter selection to study the beta-globin locus (Imam et al., 2000). Further developments include the expression of *recA* from a separate plasmid (Lalioi and Heath, 2001).

Alternatives to the *recA* recombination strategies are those that utilize the *recE/recT* pathway (equivalent to the pathway used by the λ *red* genes). These methods require much smaller regions of homology, between 40 to 60 bp. The targeting molecule is usually double-stranded linear DNA, which can be most conveniently generated entirely by PCR. Targeting is achieved in a single step, with the provision of the recombination proteins encoded from either a separate recombination plasmid or from the relevant genes that have been targeted to the *E. coli* chromosome.

The evolution of *recE/recT* dependent methods has mainly occurred within three different groups. The first group to establish a method was F. Stewart's group in EMBL, Heidelberg (Zhang et al., 1998) and in subsequent publications coined the term 'ET-cloning' to describe methods using the *recE/recT* recombination pathway. Subsequently methods evolved from P.A. Ioannou's group in The Murdoch Institute, Melbourne (Narayanan et al., 1999), using the term 'GET Recombination' and from researchers at The National Cancer Institute-Frederick, Maryland within D. Court's and N.A. Jenkins and N.G. Copeland's groups (Yu et al., 2000).

One of the aims of improving recombination mediated cloning strategies has been to optimise the expression of the genes that drive the homologous recombination reaction in an attempt to increase targeting efficiencies. A second aim was to generate strategies to control the 'window of recombination activity' to limit the undesirable recombination of the target molecule at locations other than the target. These developments are discussed in detail below.

The discovery of the RecET pathway originated from work by Clark and co-workers who found evidence for multiple recombination pathways in *E. coli* (for review see Clark and Sandler, 1994). By blocking the RecBCD pathway in *E. coli* by producing *recBC* mutants it was possible to identify further mutations that could induce the *recE* or *recF* pathways thereby restoring recombination proficiency.

The *sbcA* mutation is known to suppress the *recBC* phenotype by activating the expression of the *recE/recT* operon, encoded by part of the λ prophage integrated into *E. coli* K12 strains (Clark and Sandler, 1994; Clark et al., 1984; Hall et al., 1993). In order to use the *recET* pathway to perform homologous recombination in any *E. coli* strain, a plasmid was developed that could recapitulate the recombination status of *sbcA recBC* mutants (Zhang et al., 1998). This plasmid was pBAD-ET γ , it has a truncated *recE* gene under the inducible control of the L-arabinose inducible pBAD promoter (Guzman et al., 1995) and *recT* constitutively driven from the EM7 promoter. The *red γ* gene in this plasmid was driven by the Tn5 promoter to recapitulate *recBC* status, since it inhibits the action of the RecBCD exonuclease, thereby inhibiting the degradation of double-stranded linear fragments.

Further publications by the Stewart group sought to optimise the parameters of the ET-cloning procedure. It was determined that the λ equivalent genes for *recE* and *recT*, namely *red α* (or *exo*) and *red β* (or *bet*) respectively, could replace the function of the bacterial genes and mediate recombination with equal efficiency. The resultant plasmid was pBAD $\alpha\beta\gamma$, in which α is under the control of the inducible pBAD promoter (Guzman et al., 1995) and β driven from the EM7 promoter. As above, γ is driven by the Tn5 promoter (Muylers JP, 1999). This plasmid was applied to the modification of BACs (Muylers JP, 1999) and 'BAC Trimming' (Hill et al., 2000). By coordinating α and β under pBAD control and placing γ under the separate constitutive control of the EM7 promoter pBAD-RedGam was generated and used to drive the introduction of point mutations to BACs using a two-step procedure with *sacB-neo* second step counter selection (Muylers et al., 2000).

The construction of the pR6K116/BAD $\alpha\beta\gamma$ recombination plasmid included the introduction of the high copy R6K origin and *pir116*, which is required for its replication (Haffter et al., 1996). This plasmid was found to be more efficient at mediating DNA sub-cloning than any of the previously described plasmids (Zhang et al., 2000).

At a similar time Ioannou's group developed the pGETrec plasmid, which was derived from A.F. Stewart's plasmid pBAD24-trecET, by the addition of γ gene (Narayanan et al., 1999). pGETrec differs from pBAD-ET γ since all the genes on the former are under the inducible control of the pBAD promoter. The aim of co-ordinating their expression in this way was to control both the induction of the *recET* pathway and the inhibition of *recBCD* nuclease activity. This system has been applied to introduce mutations to BACs harbouring the β -globin locus, to recapitulate disease mutations (Nefedov et al., 2000) and to introduce GFP reporter constructs (Orford M, 2000).

A different approach than that of using a plasmid harbouring the genes necessary for recombination, was the use of *E. coli* strains engineered to harbour the genes for inducible homologous recombination on the *E. coli* chromosome. These strains have been termed in the literature as 'hyper-recombinogenic strains', since they perform homologous recombination with high efficiency and are exemplified by strains KM22, DY329 and YZ2000. KM22 was constructed by replacing the *recBCD* genes on the *E. coli* chromosome of strain AB1157, with the λ genes *red α* and *red β* , under the control of the bacterial promoter P_{lac} (Murphy, 1998). The intended use of KM22 was to facilitate the manipulation of the *E. coli* genome, as was the production of strain DY329 (Yu et al., 2000), which has the λ genes $\alpha\beta\gamma$ under the control of the temperature sensitive cI⁸⁵⁷ repressor. However, the utility of such strains for the modification of DNA molecules other than the chromosome was apparent and demonstrated by the manipulation of BAC DNA (Lee et al., 2001; Swaminathan et al., 2001). In addition, the combination of a plasmid expression system, pR6K116/BAD/recT, with the hyper-recombinogenic *E. coli* strain YZ2000, was demonstrated as capable of sub-cloning the neomycin selection marker from a background of mouse ES cell genomic DNA (Zhang et al., 2000).

The advantages of using hyper-recombinogenic strains over plasmid-based expression of the recombination genes are four-fold: the systems that use the λ recombination genes seem to be 50- to 100-fold more efficient for genes expressed

from the chromosome rather than from a plasmid; a further drug selectable marker is available to select for the introduction of mutations to the target molecule since it is not required to select for the presence of a recombination plasmid; the problem of plasmid incompatibility between the target and recombination plasmid are circumvented; and the inducible expression from the *E. coli* chromosome is less leaky than the inducible plasmid systems developed thus far (Lee et al., 2001).

In addition to the development of highly efficient *E. coli* strains that utilise the recET pathway, the other advantages that exist in using this pathway rather than the *recA* pathway and associated strategies are: that the targeting molecule may easily be generated by PCR, since the technique is efficient with 40 to 60 nucleotides of homology; and that currently protocols exist that can deliver point mutations and small deletions to the target molecule without the concomitant introduction of selection markers, thus increasing the variety of mutations that can be made and investigated.

In summary the technology exists to rapidly and efficiently make any desired modification to a BAC molecule that would facilitate the functional investigation of the coding, structural or regulatory elements contained within the cloned insert. In the context of the strategy developed within this thesis, the use of a BAC vector permitted the structural and regulatory elements of the alpha-globin locus to be encompassed in a single vector. Additionally, the use of the BAC presents the opportunity to use recombination mediated cloning strategies to flexibly manipulate the alpha-globin locus in an unprecedented manner.

1.5 The use of site-specific recombination systems for genome manipulation.

Typically site-specific recombination (SSR) is mediated between two specific sequences of DNA, known as recombination sites. The SSR reaction is catalysed by a recombinase protein that recognizes these sites and mediates the exchange exclusive of any other proteins or ATP. SSR systems of this kind include *Cre-loxP* (Abremski and Hoess, 1984), *FLP-*frt** (McLeod et al., 1986) and the less commonly applied λ *int* system (Nash et al., 1987).

The Cre-*lox* system originates from the P1 bacteriophage and its functions within the phage are to: circularise the phage following infection; monomerise plasmid dimers during lysogenic phage replication; and integrate the phage genome to the bacterial chromosome site *loxB* (Hoess et al., 1982). It is therefore surprising that this system, which has evolved to manipulate the genome of a phage, is able to manipulate megabases of eukaryotic genome and effect chromosome engineering (as discussed below). However, to date the most widely applied SSR system for mammalian genome manipulation has been Cre-*lox*.

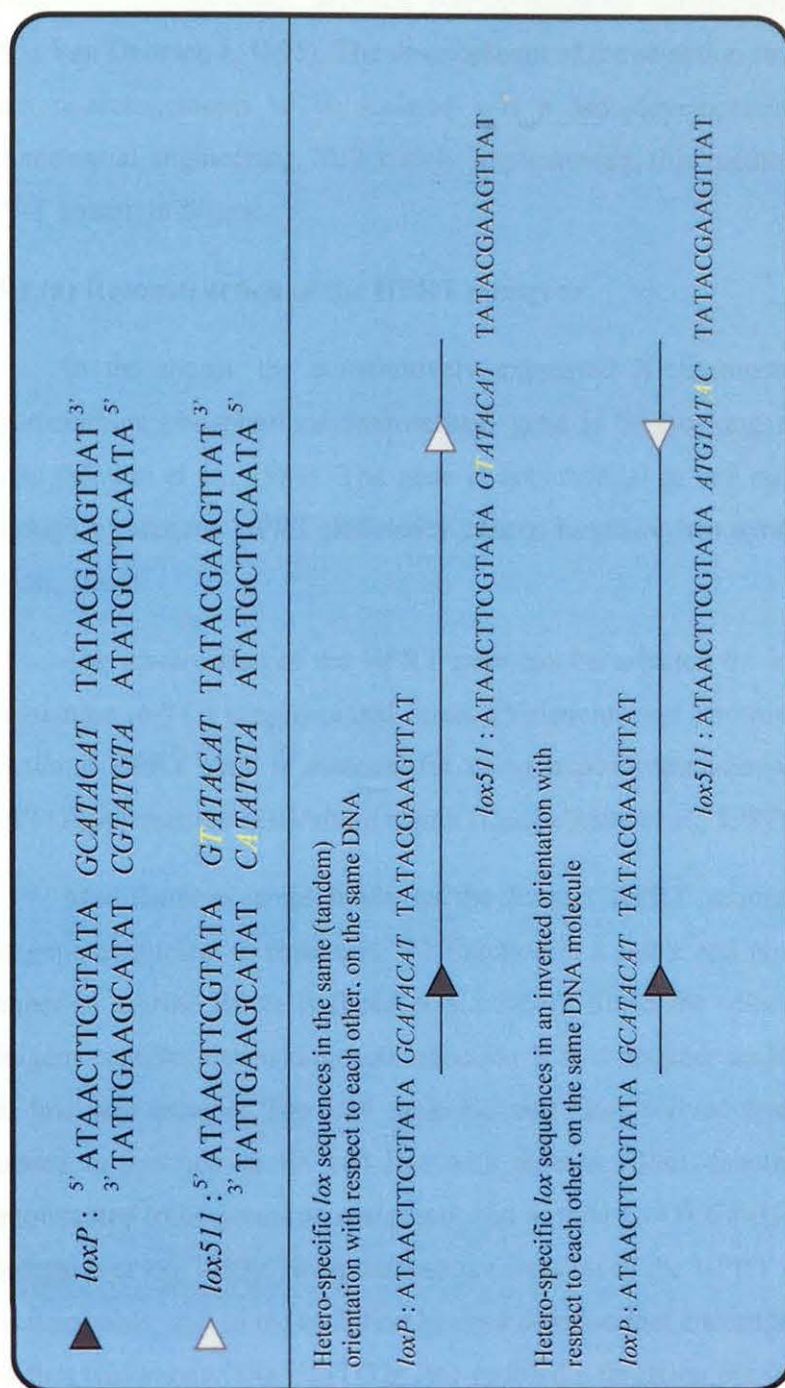
Cre is a 38 kDa recombinase and its recognition site is a 34 bp sequence, termed *loxP*. The central 8 bp of the *lox* sequence is known as the spacer region and gives the *lox* site directionality, as shown in Fig. 1.1 (Hoess et al., 1986). The remainder of the *lox* sequence forms two 13 bp inverted repeats at either side of the spacer region. Cre recombinase catalyses site-specific recombination (SSR) between *lox* sites of identical sequence (Sternberg, 1981; Sternberg and Hamilton, 1981; Sternberg et al., 1981). The variety of manipulations that can be performed by SSR are shown in Fig. 1.2. As a point of nomenclature, flanking sequences with *lox* sites that have the same orientation is commonly referred to as ‘*floxing*’ (with *frr* sites it is known as ‘*flirting*’), again see Fig. 1.2.

Cre-*lox* has been diversely applied to manipulate eukaryotic genomes. The largest manipulations made with Cre-*lox* have been on the scale of chromosomal translocations, deletions and inversions of megabases of DNA, as discussed below. The initial applications of Cre-*lox* however, manipulated smaller regions of DNA, as discussed in Section 1.5.2.

1.5.1 Chromosome engineering

Chromosome engineering is possible by combining gene-targeting with site-specific recombination (SSR) technology. The potential of Cre-*lox* technology to be able to manipulate large regions of a eukaryotic genome *in vivo* was first established when the deletion, inversion and duplication of a 3 cM region in ES cells was demonstrated (Ramirez-Solis R, 1995). In a step-wise process, *loxP* sites

Fig. 1.1 The sequence of *loxP* and *lox511* and their diagrammatic representation within this thesis



The sequence and the *loxP* and *lox511* sequences are shown in Fig. 1.1. Additionally the diagrammatic representation of the *lox* sites is also shown. The *lox* sites are made up of two 13 bp inverted repeats which bound a central 8 bp spacer region, which is italicised. As illustrated the *loxP* and *lox511* sequences differ by a single base pair within the 8 bp spacer region, which is highlighted in yellow in the *lox511* site. This difference in their sequence is the reason that these two sites will not participate in a Cre-mediated SSR reaction with each other. As shown, the spacer region also confers directionality to the *lox* sequences due to its lack of symmetry. This is in contrast with the inverted repeats, which appear to have the same sequence in either orientation of the *lox* sites.

were sequentially introduced on separate gene-targeting vectors and then transient expression of Cre from an expression cassette mediated the deletion. Since then, similar Cre-*lox* engineered chromosomal manipulations have been reported, in addition to translocations between non-homologous chromosomes (Smith et al., 1995; Van Deursen J, 1995). The development of the selection strategy that enabled these re-arrangements to be isolated was a key development in the field of chromosomal engineering. Still widely implemented, this method reconstructs the HPRT human minigene.

1.5.1 (a) Reconstruction of the HPRT minigene

In the mouse the constitutively expressed X-chromosome-linked HPRT (hypoxanthine phosphoribosyltransferase) gene is 33 kb long and contains nine exons (Melton et al., 1984). The gene is nonessential in cell culture and in mice, although in humans HPRT deficiency causes Lesch-Nyhan syndrome (Lesch and Nyhan, 1964).

The inactivation of the HPRT gene can be selected for in ES cells with 6-thioguanine (6-TG) supplemented media (Valancius and Smithies, 1991), whilst a functional HPRT gene is selected for using hypoxanthine-aminopterin-thymidine (HAT) supplemented cell culture media (Doetschman et al., 1987).

Modifications were made to the human HPRT sequence to produce a minigene competent to restore HPRT function in a stable and position independent manner in murine ES cells (Reid et al., 1990). Since the selection of the HPRT minigene requires that endogenous function is first ablated, an HPRT-deficient ES cell line was created. The E14 male ES cell line, derived from 129 mice, was mutated to produce an ES cell line with ablated HPRT function. This line was demonstrated to be germline competent and termed E14TG2a (Hooper et al., 1987; Thompson et al., 1989). No spontaneous reversion of the HPRT ablated phenotype was detectable, due to the mutation being a deletion that encompasses the promoter and first two exons. The E14TG2a line enabled a targeting construct containing an HPRT minigene to be selected for without the necessity to first ablate the

Fig. 1.2 Some of the genetic manipulations possible using SSR systems

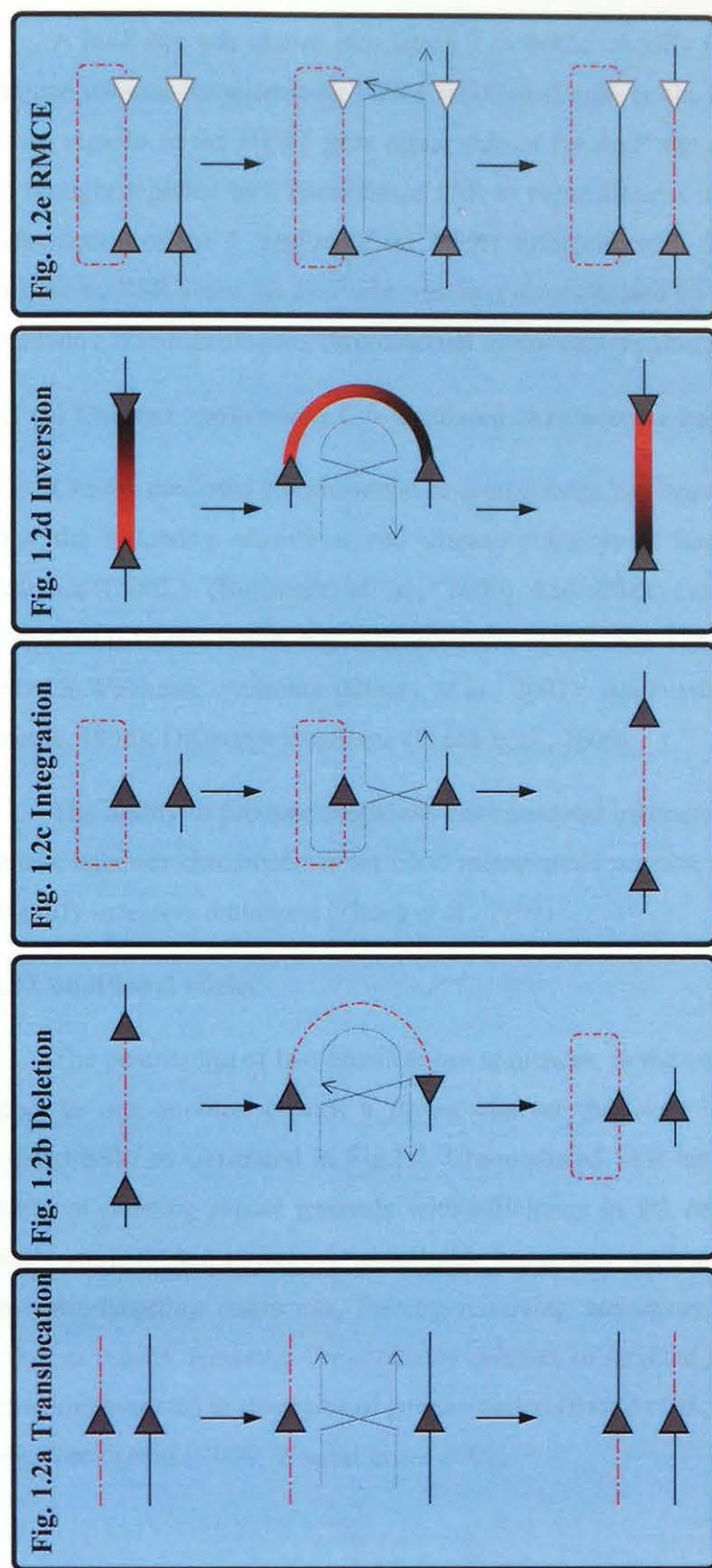


Fig. 1.2a illustrates the interaction between two lox sites on separate linear DNA molecules to result in a translocation. Fig. 1.2b shows the deletion of a floxed interval between two identical lox sites located on the same linear molecule, i.e. *in cis*, in direct orientation with respect to each other. Fig. 1.2c illustrates SSR to mediate the integration of a circular molecule into a linear molecule, in essence this is the reverse of the deletion reaction. Fig. 1.2d demonstrates that SSR between identical lox sites that are orientated in an inverted orientation causes inversion of the floxed interval, rather than its deletion. Fig. 1.2e illustrates RMCE. RMCE necessitates the use of non-compatible or hetero-specific lox sites, e.g. *loxP* and *lox511*. As shown, the floxed interval in the circular molecule is exchanged with that of the linear molecule in a directional manner, due to the exclusive SSR between compatible lox sites. An inverted orientation of the hetero-specific lox sites is shown, however since their SSR reactions are exclusive both direct and inverted orientations may be implemented.

endogenous HPRT function (Selfridge et al., 1992).

A *loxP* site was cloned into intron 3 to further modify the Reid *et al*, 1990 minigene without compromising HPRT function (Smith et al., 1995). This enabled the two regions of the HPRT gene either side of the *loxP* site to be separated and then brought together by Cre-mediated SSR to reconstitute a functional gene. The reconstruction of the 5' region of the HPRT minigene with the 3' region of the minigene by SSR about the *loxP* site was first demonstrated by Smith *et al* in 1995 to mediate a non-homologous chromosomal translocation (Smith et al., 1995).

1.5.1 (b) Current applications Cre-mediated chromosome engineering

Cre-*lox* mediated chromosomal re-arrangement has been used effectively to mimic the following conditions and disease phenotypes: human acute myeloid leukaemia (AML) (Buchholz et al., 2000) and other cancers arising from chromosomal translocation, e.g. lymphoblastic leukaemias (Rabbitts et al., 2001); Beckwith-Wiedman syndrome (Cleary et al., 2001); synpolydactyly (Zakany and Duboule, 1996); DiGeorge syndrome (Puech et al., 2000).

The ability to produce megabase chromosomal inversions has been applied to create balancer chromosomes for ENU mutagenesis screens, the aim of which is to identify recessive mutations (Zheng et al., 1999).

1.5.2 Conditional alleles

The positioning of two identical *lox* sequences, in the same orientation with respect to one another, creates a *floxed* interval that may be deleted by Cre-mediated SSR, as illustrated in Fig.1.2. Cre-mediated SSR has been shown to be capable of deleting *floxed* intervals with efficiency in ES cells. This facet was initially implemented to remove undesirable sequences, such as selection markers, from gene-targeting constructs, thereby removing sequences that may generate artefactual results. However Cre-mediated deletion of targeted sequences has been further implemented in strategies of gene mutation (Brault et al., 2001; Chang et al., 1999; Tomita et al., 2000; Wagner et al., 2000).

In order to understand gene function one avenue of investigation is to ablate the function of a gene by deleting its sequence either in part or entirely from the genome and then study the resulting phenotype. The organism that results from such mutations is commonly referred to as a 'knock-out'.

Knock-out mice may be created by removing a number of exons critical to the function of the gene by gene-targeting. However, the assessment of the knock-out phenotype from such mutations should also include a rigorous assessment of its genome, to ensure that the remaining gene sequences do not provide any residual function. To conclusively remove any residual gene function one solution would be to create a deletion of the entire gene. In some cases this may be impractical with gene-targeting methods due to the size of the gene and achieved more easily with *Cre-lox* technology. However, the deletion of entire genes may concomitantly remove sequences that affect the expression of other genes.

A major obstacle that may be encountered during the functional assessment of a knock-out phenotype is lethality due to the mutation, present in every cell of the embryo, disrupting the normal developmental path of the organism. This presents a problem if the aim of the investigation is to study the consequences of ablated gene function in the adult organism alone.

To circumvent this problem, strategies were developed that could restrict the mutation of a gene in a spatio-temporal manner to a population of cells in which its expression had functional relevance. The ability to restrict ablation or mutagenesis of a gene in a spatio-temporal manner was achieved by the development of conditional alleles. Conditional alleles can be generated by gene-targeting in ES cells and include genes whose sequences are either entirely *floxed*, or genes that have had critical exons *floxed*. These ES cells can then be used to generate transgenic mice that should not be functionally compromised and so develop into adult mice harbouring the *floxed* allele.

In parallel, transgenic mouse lines were developed in which the expression of Cre recombinase may be controlled in a spatio-temporal manner, due to the expression of Cre from a suitably restricted promoter. The promoter may be

restricted due to its expression being exclusive to a specific tissue type (Gu et al., 1994) or to a stage of development or both (<http://www.mshri.on.ca/nagy/cre.htm>). In addition, fusion of the promoter to tet operator sequences (St-Onge et al., 1996) or fusion of Cre recombinase to hormone receptor domains (Kellendonk et al., 1996; Zhang Y, 1996), permits the temporal induction of the recombinase. As an alternative to transgenic lines, Cre may also be delivered by injection of the relevant tissue with an adenovirus vector carrying Cre (Wang et al., 1996).

The inter-cross of a transgenic Cre mouse with a mouse harbouring a *floxed* allele should generate double-mutant progeny. The double-mutants undergo SSR, following Cre expression or induction, to delete the *floxed* interval thereby deleting either a gene or critical exons, to cause the mutation. The phenotype of the resulting mice may be then analysed. The use of conditional alleles is implemented most advantageously to circumvent the problem of embryonic lethality, due to the mutation of a gene critical to development. Thus conditional alleles permit the investigation of a gene's function in the relevant tissue within an adult model organism.

1.5.3 Recombinase Mediated Cassette Exchange (RMCE)

Deletion of conditional alleles is achieved by *floxing* sequences with compatible *lox* sites, in direct orientation with respect to each other, as illustrated in Fig. 1.2 and outlined above. In contrast, the use of variant or hetero-specific *lox* sites, permits the directional exchange of a similarly *floxed* interval, see Fig. 1.2. The common term for such a reaction is recombinase mediated cassette exchange (RMCE).

Recently, Cre-*lox* and FLP-*frt* recombinase systems have been applied to mediate the directional exchange of small (1-2 kb) cassettes of DNA in strategies that share the same rationale and have been termed recombinase mediated cassette exchange (RMCE) (Bouhassira et al., 1997), segmental genome replacement (Bethke and Sauer, 1997) or double-reciprocal cross-over (Bode 1997). The rationale behind these experiments is summarised in Fig. 1.2 and for convenience will be referred to as RMCE.

RMCE involves the use of a SSR system, FLP-*frt* and Cre-*lox* have been cited in literature thus far (Bode et al., 2000). SSR sites are used to flank the interval to be exchanged as shown in Fig.1.2. The recombination sites are hetero-specific in order to promote inter- rather than intra-molecular SSR. RMCE was first demonstrated using the FLP-*frt* SSR system, using wild-type *frt* and a hetero-specific *frt* generated by mutation (Schlake and Bode, 1994), and since that time the published use of the FLP-*frt* system has been by Bode and colleagues. The Cre-*lox* system was initially found to be more suitable for cells in culture due to its thermostability and may explain the limited application of the FLP-*frt* system thus far, although more thermostable FLP recombinase variants have become available (Buchholz et al., 1998; Rodriguez et al., 2000). Since the initial 1994 publication two further reports from Bode and colleagues have demonstrated the application of the FLP-*frt* RMCE system, most notably in 1998 when RMCE was first applied in ES cells (Seibler et al., 1998).

In 1997 two labs independently reported the use of different RMCE strategies utilizing Cre-*lox*, with *loxP* and *lox511* sites. The term 'RMCE' was coined by Bouhassira and was implemented with a 'plug and socket' method of selection (Bouhassira et al., 1997). In the same year Sauer used a different selection strategy to implement 'segmental genome replacement' and in 1999 was the first to implement RMCE in ES cells using the Cre-*lox* system and tested a variety of selective procedures (Soukharev et al., 1999)

The majority of reported RMCE experiments use the Cre-*lox* SSR system, with *loxP* and *lox511* (Hoess et al., 1986) being the hetero-specific sites of choice. However the low incidence of SSR between *loxP* and *lox511* in some cases has led to the use of another *lox* variant, *lox2722* (Kolb, 2001), or more recently the *lox66* and *lox72* pair (Zhang and Lutz, 2002). The research group headed by Bouhassira was the first to report the use *loxP* and *lox511* in a tandem arrangement (Bouhassira et al., 1997), but then altered their strategy to arrange the sites in an inverted orientation (Feng et al., 1999). The inverted orientation was chosen in order to eliminate the possibility of deleting the interval between the hetero-specific *lox* sites

should SSR between them occur (see Fig.1.2), thereby, in theory, increasing the likelihood of generating an RMCE event.

The efficiency of the RMCE reaction is reported by most studies to be high and affected by the selection strategy employed. However, the application of RMCE between studies has little similarity and so a comparison of their efficiencies has little relevance.

Sauer was the first to use a Cre-*lox* strategy to introduce DNA into a pre-designated site in a complex mammalian genome by SSR (Sauer and Henderson, 1988) (see previous discussion). Sauer then went on to compare the efficiencies of using either an RMCE replacement strategy or single *lox* site insertion approach. The RMCE strategy was found to be 20-fold more efficient than the equivalent system with just a single *lox* site and furthermore the RMCE approach was shown to be more efficient than the level of illegitimate recombination (Bethke 1997). The study used two rounds of selection, the first selection identified those cells that had taken up the RMCE construct, the second the RMCE event itself, based upon selection markers within the RMCE construct. The results were that from populations of cells that had taken up the RMCE DNA construct, 2% to 10% underwent the desired RMCE event. In a separate investigation, it was demonstrated that RMCE events could be identified from populations of cells shown to have taken up the RMCE construct by screening alone (Soukharev et al., 1999). This study further exemplified the efficiency of the RMCE reaction and demonstrated that selection markers need not be incorporated within the RMCE construct. However, when two-stage selection strategies are used in conjunction with an RMCE approach, for example PNS, three different research groups have obtained RMCE efficiencies of 100% (Feng et al., 1999; Seibler et al., 1998; Soukharev et al., 1999).

The application of RMCE to study a biological system has been most prolifically reported by a research team headed by Bouhassira. This group utilizes the plug and socket RMCE approach in order to investigate the genes and chromatin regulatory elements, associated with the alpha (Esperet et al., 2000) and

beta-globin loci (Walters et al., 1999). Thus far, the other biological application of an RMCE strategy has been to direct the insertion of transgenes to transcriptionally permissive loci (Kolb, 2001).

Thus far, there has been no published report describing RMCE of *floxed* or *flirted* intervals of more than 5 kb.

A recent advance in Cre-*lox* technology applied directed molecular evolution to alter the specificity of the Cre recombinase to produce a panel of novel enzymes with corresponding unique recognition sites with very little cross-reactivity to the wild-type Cre-*loxP/lox511* system (Buchholz and Stewart, 2001). The new Cre-*lox* variant systems could extend the use of RMCE strategies, for example in order to replace loci that are located on different chromosomes or on the same chromosome at distances that necessitate the use of two independent RMCE events.

In summary, the Cre-*lox* system has been demonstrated to have the capacity to effectively and flexibly manipulate a large size range of DNA, from a complex and organized genome (for a review see Yu and Bradley, 2001).

1.5.4 Large interval RMCE

In essence the strategy developed within this thesis attempts to perform RMCE using a larger interval between the hetero-specific *lox* sites. For this reason a consideration of the efficiencies from the RMCE experiments reported thus far, may be informative to predictions on the efficiency of mediating a larger exchange. However, the mechanism by which a small interval RMCE event occurs may be different to the progression of a large interval RMCE, in which case, a comparison of efficiencies would be of little value.

Thus far, there have been some considerations on the factors that affect the efficiency of Cre-*lox* mediated eukaryotic genomic manipulations (Zheng et al., 2000). One possible explanation for the reduced efficiencies that are encountered with some chromosomal engineering strategies, may be the existence of higher orders of nuclear organization in eukaryotic cells. The organisation of the

chromosomal environment harbouring the *lox* sites may restrict its availability to the Cre recombinase or restrict its participation in SSR by restricting the motility and flexibility of the DNA (Vooijs et al., 2001). It was therefore necessary, in the context of this strategy, to determine whether the efficiency of the desired replacement event would be so low as to be unrecoverable, or whether the large interval RMCE event would occur with efficiency comparable to those encountered with chromosomal engineering strategies using equivalent intervals.

1.5.4 (a) The strategy developed in this thesis for large interval RMCE

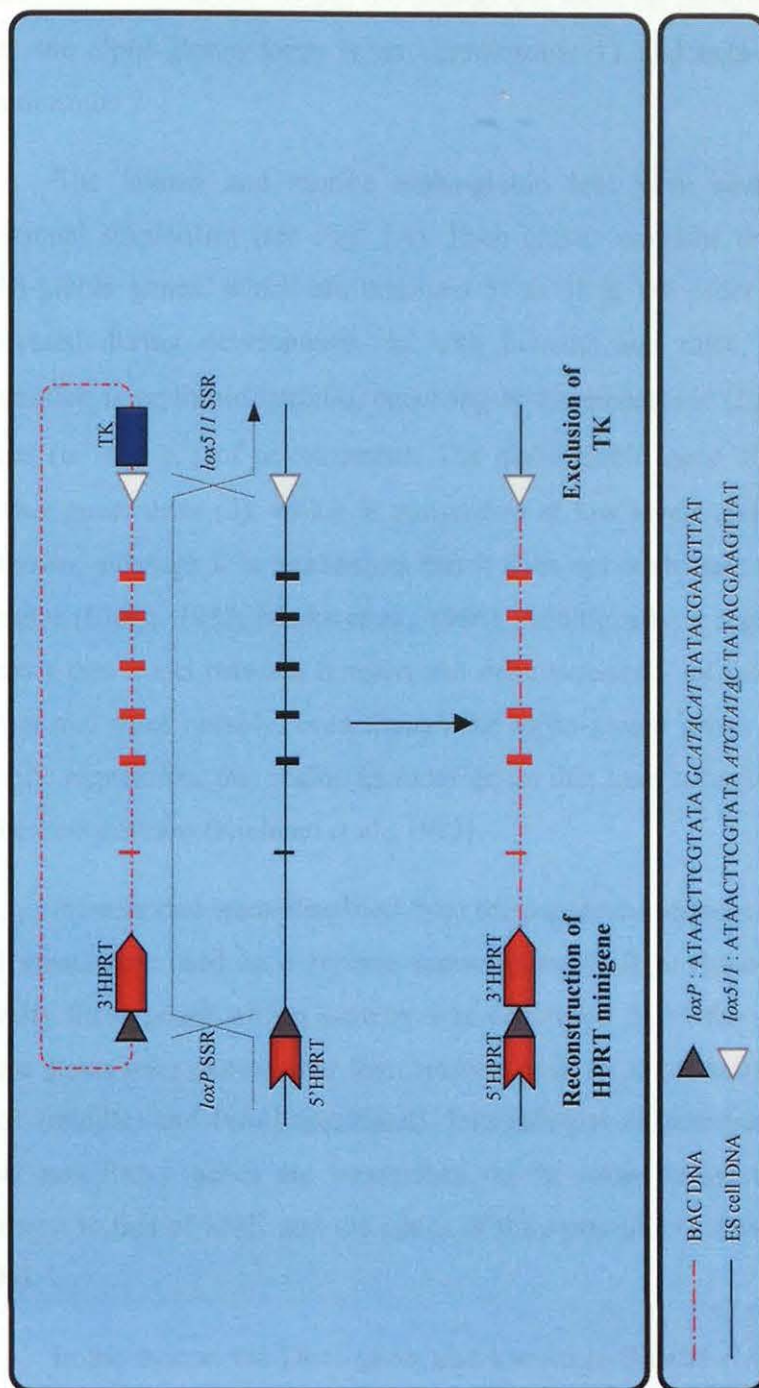
The strategy developed within this thesis combines technology from chromosomal engineering studies and large vector transgenesis, in order to develop technology to study functional genome elements in their normal genomic context. Gene-targeting and Cre-mediated SSR, were combined with the BAC vector system in a large interval RMCE strategy. The development of this strategy serves as a preliminary investigation towards the segmental replacement of the ES cell genome. An outline of the strategy for achieving the large interval RMCE event is illustrated in Fig. 1.3.

This thesis describes the progress on developing the large interval RMCE strategy, using the alpha-globin locus as a test region. The strategy utilizes Cre-mediated SSR between hetero-specific *lox* sites targeted to the ES cell genome and the genomic insert cloned into a BAC. Therefore, with reference to Fig. 1.3, the circular molecule represents the participation of the BAC, whilst the linear molecule represents the ES cell locus. As shown in Fig. 1.3, at one end of the interval, reconstruction of HPRT occurs around the *loxP* sites and at the other end the thymidine kinase is excluded from the genome upon *lox511* SSR. The execution of the desired RMCE event should therefore be possible by selecting the ES cells on the basis of their resistance to HAT and ganciclovir supplemented media.

1.6 The alpha-globin locus as a test locus

Haemaglobin is the protein tetramer that carries oxygen in mammalian arterial erythrocytes and it is comprised of two alpha-globin and two beta-globin

Fig. 1.3 The strategy for large interval RMCE in outline



As illustrated the BAC harbours the locus of interest, which has been floxed with hetero-specific *lox* sites, *loxP* and *lox511*. The locus of interest is similarly floxed in the ES cell genome. SSR around the *loxP* site should bring together the 5' and 3' regions of the HPRT minigene, resulting in its reconstruction and ability to function. Therefore cells that have undergone SSR around *loxP* should be resistant to HAT selection. The desirable SSR around the *lox511* site should exclude the TK selection marker from the ES cell genome and this event may be selected for in GANC supplemented culture media. The occurrence of the desirable large interval RMCE is therefore identified in cells that are resistant to both HAT and GANC supplemented culture media.

protein subunits. The genes that encode the alpha and beta globin proteins are each clustered at spatially distinct loci, an arrangement that is thought to be the result of the duplication from a common ancestral locus. In humans the alpha-globin locus is located on chromosome 16 and the beta-globin locus on chromosome 11, whilst in mice the alpha-globin locus is on chromosome 11 and beta-globin locus is on chromosome 7.

The human and murine alpha-globin loci have several structural and functional similarities (see Fig. 1.4). Each cluster contains the same number of alpha-globin genes, which are arranged 5' to 3' in the order in which they are expressed during development. In both humans and mice, alpha-globin gene expression is erythroid-specific, occurring at the embryonic (ζ), or fetal and adult stages (α^1 and α^2) of development. The alpha-globin gene cluster also includes another gene, theta (θ), which is transcribed at low levels and whose function is unknown, although it is established that it does not contribute to the haemoglobin tetramer (Clegg, 1987; Marks et al., 1986). Additionally, a region of considerable synteny that exists between humans and mice extends 5' of the alpha-globin gene cluster and most notably, even though the alpha-globin genes have an erythroid-specific expression, this region includes genes that have constitutive and extensive expression patterns (Kielman et al., 1993).

Primers that were identified from the conserved sequences between humans and mice were used in a reverse transcriptase PCR analysis and were able to identify three genes whose synteny was conserved 5' of the alpha-globin locus. These genes were named after their relative position to the locus as: Dist1 (distal), Mid1 (middle) and Prox1 (proximal). Interestingly, in both humans and mice the Dist1 and Prox1 genes are transcribed on the same strand but in an opposite direction to that of Mid1 and the genes of the alpha-globin cluster (Kielman et al., 1996).

In the mouse, the Dist1 gene, also known as Rhbdf1 (MGI: 104328), has a cDNA sequence which suggests it has receptor activity and it is expressed strongly in epithelial rich tissues (Vickers et al., 1993; Kielman et al., 1996). The Mid1 gene

has been characterised and found to encode the enzyme N-methylpurine-DNA glycosylase and as such was renamed Mpg (MGI: 97073) (Engelward et al., 1993; Kielman et al., 1995), it is expressed at high levels in the testis. The Prox1 gene, also known as Phg (MGI: 84809) has a constitutive expression pattern and is therefore thought to encode a house-keeping gene, although its function has not been definitively established (Kielman et al., 1996). The human homolog of Prox1, hProx1, is also known as C16orf35 or -14 (Vyas et al., 1995).

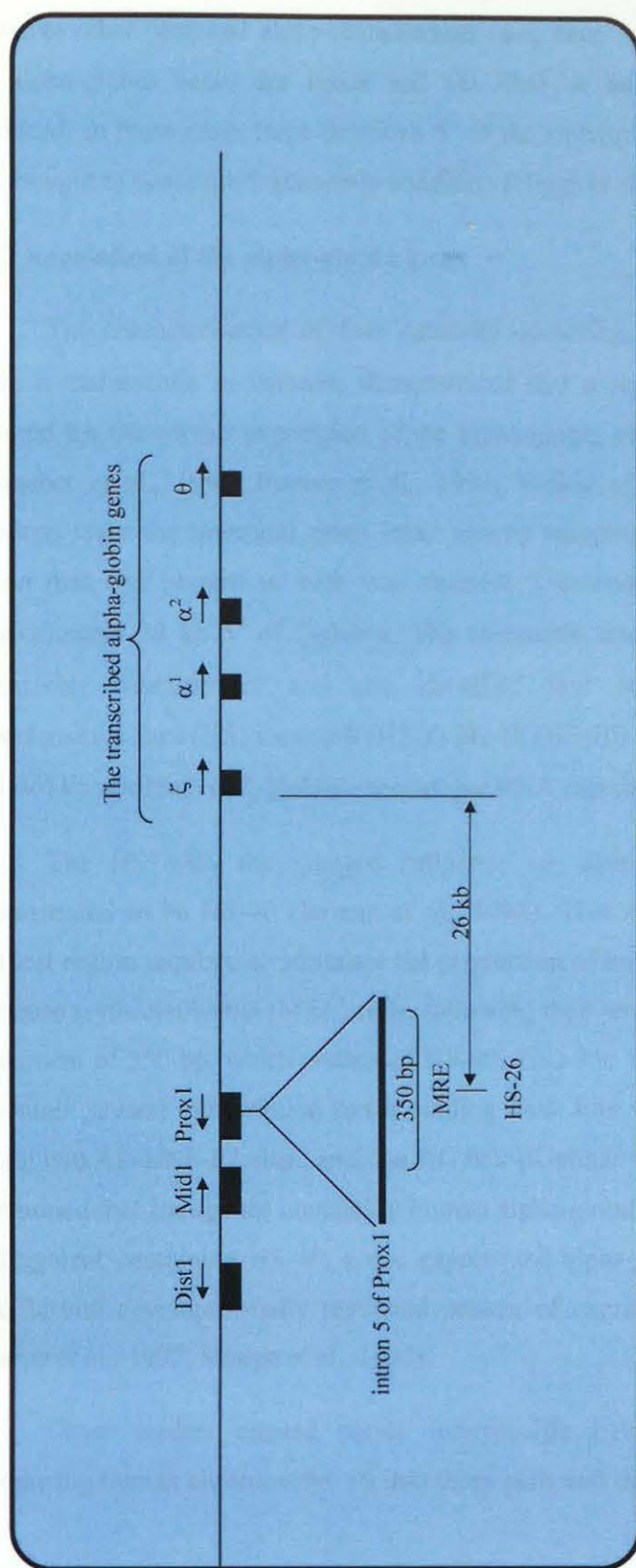
The conservation of sequences 3' of the alpha-globin locus between humans and mice is less apparent, with the only identifiable conserved sequences being repetitive elements related to retroviral sequences. In humans, the first gene encountered 3' of the alpha-globin locus is located at a distance of 11.5 kb and is *LUC7L*. However in the mouse the orthologue of *LUC7L* is found on chromosome 17. Thus far, the identification of putative regulatory sequences within this 3' region has been unsuccessful (Flint et al., 2001).

1.6.1 Diseases associated with the alpha-globin locus

Under normal circumstances transcription from the alpha-globin locus is coordinated with that of the beta-globin locus to generate a normal haemoglobin tetramer. However, there are various anaemias that occur due the disproportionate amount of either alpha- or beta-globin protein sub-units, which leads to the formation of abnormal haemoglobin proteins that have an inability or inefficiency to either associate or dis-associate from oxygen. These are the thalassemias. Alpha-thalassemias are characterised by an abundance of beta-globin proteins, the result of which is the formation of unstable haemoglobin tetramers, which have a decreased efficiency for oxygen transport. This form of anaemia is referred to as microcytic anaemia (in contrast to macrocytic anaemia e.g. sickle-cell anaemia), since the erythrocytes are visibly normal despite their reduced functional competency (Higgs et al., 1989).

The abundance of beta-globin proteins in alpha-thalassemias may be due to the over-expression of the beta-globin genes or (more commonly) a deficiency in the amount of protein generated from the alpha-globin genes, which may be caused

Fig. 1.4 Arrangement of the murine alpha-globin locus and genes 5' of the locus



The arrangement of the murine alpha-globin locus on chromosome 11 is as shown (not to scale). The four transcribed alpha-globin genes are represented with the symbols ζ , α^1 , α^2 and θ . Approximately 20 kb 5' of the ζ gene in the murine locus are three genes whose positions are conserved between humans and mice, these are Distal (Dist1), Middle (Mid1) and Proximal (Prox1). All of the genes shown above are transcribed from the same strand of DNA, the direction of their transcription is indicated by an arrow. In humans a *DNAse* 1 hyper-sensitive site (HS) was found within intron 5 of human Prox1 and this genetically sign-posted a 350 bp region, termed the Major Regulatory Element (MRE), which was shown to be required for the correct expression of the human alpha-globin genes. The homologous sequences to the human MRE in the mouse also reside in intron 5 of the murine Prox1 gene and this intron also contains a *DNAse* 1 HS site, which is 26 kb 5' of ζ and therefore termed HS-26.

by a mutation or deletion of one of the genes from the alpha-globin cluster. However other inherited alpha-thalassemias have been presented for which all of the alpha-globin genes are intact and yet little or no alpha-globin protein is produced. In these cases large deletions 5' of the alpha-globin locus were detected and thought to cause the thalassemic condition (Higgs et al., 1989).

1.6.2 Regulation of the alpha-globin locus

The characterisation of four naturally occurring, inherited deletions, that cause α -thalassemia in humans, demonstrated that a region 5' of the cluster is required for the correct expression of the alpha-globin genes (Hatton et al., 1990; Liebhaber et al., 1990; Romao et al., 1991; Wilkie et al., 1990). All of these deletions leave the structural genes intact and by mapping the deletions a minimal region that was present in each was defined, it extended for 20.4 kb and was approximately 30 kb 5' of ζ -globin. The chromatin structure in this region was extensively characterised and this identified four erythroid-specific DNaseI hypersensitive sites (HS) located 8 (HS-8) kb, 10 (HS-10) kb, 33 (HS-33) kb and 40 (HS-40) kb upstream of ζ -globin's messenger RNA cap site (Higgs et al., 1990).

The HS with the greatest influence on alpha-globin expression was demonstrated to be HS-40 (Jarman et al., 1991). This study determined that the smallest region required to stimulate the production of human alpha-globin mRNA in mouse erythroleukemia (MEL) cells, following their terminal differentiation, was a fragment of 350 bp, which contained HS-40. This 350 bp region was also shown to contain several transcription factor binding sites: four GATA-1 sites; four CAC boxes; two AP-1/NF-E2 sites; and one AG box (Kielman et al., 1994). Studies then determined that transgenes containing human alpha-globin genes linked to this 350 bp fragment containing HS-40, could express the alpha-globin genes in a tissue-specific and developmentally regulated pattern of expression in transgenic mice (Sharpe et al., 1992; Sharpe et al., 1993).

Other studies created stable interspecific hybrids in MEL cells by introducing human chromosome 16 into these cells and then assessing them for the

level of human alpha-globin gene expression. The result was that human alpha-globin genes were expressed, although at levels lower than the endogenous mouse genes (12 % to 56 % of mouse globin levels) (Higgs et al., 1998). In this same system the effect on human alpha-globin gene expression of replacing a 1 kb fragment containing HS-40 with a neomycin gene, was tested. The neomycin gene was flanked by *frt* sites, which enabled its deletion following FLP recombinase expression. MEL cells with either the neomycin present or FLP-deleted, from human chromosome 16, demonstrated a severe down-regulation in the expression of human alpha-globin genes (Bernet et al., 1994).

Therefore, the conclusion from these studies was that HS-40 was the element necessary and sufficient to regulate human alpha-globin expression.

The investigation of the murine alpha-globin sequence, with a homology-based search, was used to locate the murine equivalent of HS-40. This investigation discovered sequence homology 26 kb 5' of murine ζ -globin and this site co-localised with a HS and as such was termed HS-26 (Kielman et al., 1994). The functional equivalence of HS-40 and HS-26 was further suggested by their conserved location between humans and mice within intron 5 of the *Prox1* gene (Kielman et al., 1996).

HS-26 co-localises with several transcription factor binding sites conserved between HS-40 and HS-26, these include: three GATA-1 sites; one CAC box; two AP-1/NF-E2 sites; and one AG box (Kielman et al., 1994). In comparison with HS-40, this is one less GATA-1 and three less CAC box transcription factor binding sites.

To determine whether HS-40 and HS-26 were functionally equivalent, a 350 bp fragment containing HS-26 was linked to the globin genes of the human locus and then introduced to MEL cells. This investigation determined that HS-26 was a much weaker enhancer for human-globin genes than HS-40, however this study suggested that this could be due to an incompatibility between the murine and human sequences, rather than an inability of HS-26 to act as an enhancer

(Bouhassira et al., 1997). The same study investigated the effects of replacing the 350 bp region containing HS-26, with a neomycin gene and this was achieved by gene-targeting in ES cells. The effect of this mutation was assessed in murine chimeras (note, no report of germline transmission was made in this study). Each chimeric mouse harboured a population of erythrocytes that appeared thalassemic, which suggested that HS-26 is required for alpha-globin regulation. However it was not possible to truly establish the extent to which HS-26 controls globin expression since the level of chimerism was an unknown in this study. Furthermore, the effects of the neomycin gene, located within intron 5 of Prox1, were not established, neither on the regulation of the alpha-globin genes nor on the expression of Prox1 (Bouhassira et al., 1997).

Most of the evidence, as outlined above, suggested a functional equivalence between HS-40 and HS-26. However, since the knock-out of HS-26 had not generated conclusive data, a more definitive study on the effect of deleting HS-26 on alpha-globin expression was undertaken. HS-26 and the transcription factor binding sites associated with it were replaced with a *floxed* neomycin gene, a replacement of 1.3 kb. This study was anticipated to be equivalent to that of Bernet et al., 1994, in which the HS-40 was deleted with a resulting down-regulation of human alpha-globin gene expression. However, upon deletion by SSR of the *floxed* neomycin gene (that replaced HS-26) the anticipated elimination of murine alpha-globin gene transcription did not occur. The mRNA expression from the alpha-globin genes persisted to between 50 and 80% of normal levels (Anguita et al., 2002). This study therefore concluded that other, as yet uncharacterised elements, contribute to control the expression of the alpha-globin genes in the mouse. Furthermore, this study exemplifies the requirement for the *in vivo* verification of sequence-homology based functional assignments.

Clearly the results from the deletion of HS-40 (Bernet et al., 1994) and the deletion of HS-26 (Anguita et al., 2002) do not demonstrate the same reduction in the expression of the alpha-globin genes, for the human and mouse alpha-globin gene cluster respectively. However it should be noted that deletion of the HS-40

region is performed in MEL hybrid cells and not in human ES cells which would provide a more comparable *in vitro* study to that of the HS-26 knock-out. This experimental difference may account for the different phenotypes generated for each of the knock-outs. Alternatively, it could be that HS-40 and HS-26 are indeed not functionally equivalent. This would be consistent with the difference in their transcription factor binding site repertoire, which is probably a reflection of the different environments in which the human and mouse alpha-globin loci reside.

Differences between the human and mouse loci include: the position of the locus in the genome, the human alpha-globin locus lies very close to the telomere, whilst the mouse locus is interstitial (Higgs et al., 1989); the human alpha-globin promoters are associated with CpG islands but the mouse promoters are largely devoid of CpG islands (Flint et al., 2001); and the mouse alpha-globin promoters contain GATA-1 binding sites, but their human orthologues do not (Flint et al., 2001).

1.6.3 The application of large interval RMCE to study the alpha-globin locus

Thus far, investigations on the alpha-globin locus have not fully established the mechanism by which the alpha-globin genes are regulated during development. Furthermore, questions still remain that relate to whether the conservation of sequence homology between humans and mice, specifically within the putative regulatory region 5' of the transcribed globin gene cluster, is indicative of sequences that have a conserved function.

To summarise the studies discussed in section 1.6.2, in humans, a 350 bp region around HS-40 was demonstrated, by its removal, as essential for the normal expression of the human alpha-globin genes. In contrast, in mice, the 350 bp region around HS-26, which exhibits sequence homology to that around HS-40, does not seem to be essential for transcription of the murine alpha-globin genes. These studies could indicate that for sequences that regulate gene expression, equivalent function can not be always be inferred by the conservation of sequence homology. However, this conclusion should be tempered with the consideration that the

experimental conditions of each of these studies was not equivalent and could therefore contribute to the unexpected difference in the knock-out phenotypes.

In order to achieve a better comparison of the function of HS-40 and HS-26 further investigation of the human and mouse loci are required, preferably with studies that are performed under more comparable conditions. The large interval RMCE strategy developed within this thesis is tested in murine ES cells and therefore to have a comparable human system the human alpha-globin locus would require investigation in human ES cells. Although, murine ES cells do not express alpha-globin, upon *in vitro* erythroid differentiation the resultant cell types have been shown to express all of the alpha-globin genes (Lindenbaum and Grosveld, 1990) and this protocol for differentiation could be applied to human ES cells. The knock-out of the HS-40 region from human ES cells may achieve a more accurate report on the consequences of this mutation, since it would be studied on a more usual genomic background of human chromosomes. This is unlike the existing study which examines the knock-out phenotype in differentiated MEL cells on a genomic background of murine chromosomes (Bernet et al., 1994).

However the knock-out of HS-40 in human ES cells could be achieved by conventional gene-targeting technology and was not therefore behind the motivation to develop a strategy of large interval RMCE. It was anticipated that the benefit of developing a large interval RMCE in order to generate knock-out mutations of either HS-40 or HS-26, was the increased level of flexibility with which these putative regulatory regions could be mutated, above that possible with conventional gene-targeting.

The existing knock-out mutations of either HS-40 or HS-26 were made by replacing approximately 1 kb of sequence, which included the HS site and associated transcription factor binding sites, but may have also included other, as yet uncharacterised sequences necessary for the correct transcription of the alpha-globin genes. Additionally the knock-out studies thus far on HS-40 and HS-26 left sequences within the putative regulatory domain, which may have contributed to

the mutant phenotype, thereby complicating the analysis of the data from such studies.

Therefore the aims of the developing the large interval RMCE strategy were to enable a more precise and flexible mutational approach to investigate the alpha-globin locus, or indeed any given locus and to enable the mutational analysis of a region without the concomitant introduction of any other sequences, thereby permitting the accurate report of the intended mutation. Instead of a 1 kb deletion with a selection marker, RMCE could permit the mutation of the individual transcription factor binding sites, with mutations as small as individual base-pairs possible, without the concomitant introduction of any other sequences. Additionally, any number or combination of mutations could be rapidly made to the BAC vector by homologous recombination in the *E. coli* cell and then delivered in one SSR event to the ES cell interval, thereby expediting the investigation of the locus by removing the requirement for lengthy experiments with successive rounds of gene-targeting. For example, it would be possible to add transcription factor binding sites around HS-26 to make it equivalent to the HS-40 region in humans. If the HS-40 and HS-26 regions are indeed functionally distinct, this analysis may determine those sites that are critical to this functional difference.

Since the inception of this thesis RMCE has been demonstrated on a scale of 1 kb to 5 kb. This interval would be adequate to encompass the HS-40 or HS-26 region. However, the advantage of developing a strategy of large interval RMCE that can potentially replace 50 kb to 100 kb of sequence, are that both the regulatory region and the transcribed genes of the locus may be manipulated within a single vector and transferred to the ES cell interval in one SSR event.

For example, it may be desirable to determine the effects of a mutation within the putative regulatory region on the transcribed alpha-globin genes and this could be traced by introducing different fluorescent markers to tag the expression of each of the different alpha-globin genes. To perform such a study by conventional gene-targeting would require several rounds of gene-targeting. Alternatively it may

be desirable to mutate the putative regulatory region and co-ordinately mutate those sequences that would respond to the mutation in the transcribed genes. For example a mutation that would cause the expression of the embryonic globin gene alone could be generated and the sequences required to activate the fetal and adult genes with this mutated regulatory region could be determined.

Finally, the use of a large interval RMCE strategy would permit the exchange of the murine or human alpha-globin locus into human or murine ES cells respectively, which have been previously engineered to undergo the RMCE exchange. The ability to exchange the loci in this way would be informative on the way in which the cellular environment may affect any mutation made to the locus.

The application of large interval RMCE to study alpha-globin gene expression therefore has a number of advantages over conventional genomic manipulation strategies.

1.7 Thesis outline

The results chapters contained within this thesis describe the establishment of the desired large interval RMCE event in murine ES cells. Chapter 2 describes the establishment of a double-targeted ES cell line by the sequential targeting of two gene-targeting vectors containing sequences that mediate and select for the large interval RMCE event. Chapter 3 describes the characterisation of the double-targeted ES cell lines, in order to determine whether each vector is targeted to the same allele of the alpha-globin locus, on chromosome 11. Additionally, Chapter 3 describes the functional assessment of the *lox* sites. Chapter 4 describes the targeting of the BAC, harbouring the alpha-globin locus, by recombination mediated cloning, in order to introduce sequences that enable it to participate in the desired large interval RMCE event. Finally, Chapter 5 reports the experimental results of the RMCE experiment and Chapter 6 discusses the relevance of these results.

Chapter 2

EMBRYONIC STEM CELL TARGETING

2.1 Introduction

This chapter describes the generation of a murine ES cell line capable of participating in the large interval RMCE strategy, developed within this thesis. This was achieved by two successive gene-targeting events in the E14Tg2a ES cell line. The murine alpha-globin locus was chosen as a test locus. Collaborators, Doug Higgs and Christina Tufarelli, supplied the DNA and complete sequence data of pBAC14567, which consist of the BAC cloning vector pBeloBAC11 and a *HindIII* cloned insert of approximately 70 kb of murine 129 genomic DNA that encompassed the entire alpha-globin locus. The cloned insert in pBAC14567 was therefore isogenic with the DNA of the ES cell line that was targeted. The availability of sequence data facilitated the design of targeting constructs and the subsequent screening strategies that established a targeted event.

The murine alpha-globin locus consists of genes, pseudogenes and sequences that are necessary to regulate gene expression from the locus. The most influential of these regulatory sequences are thought to be clustered together in a region the Major Regulatory Element (MRE), which is marked by a *DNase* I hypersensitive site, termed HS-26 (Kielman et al., 1994). The MRE and structural genes of the alpha-globin locus span a genomic region of approximately 70 kb.

To mediate the RMCE of the entire locus, targeting of a *loxP* and a *lox511* site, at positions at either end of the locus, were necessary. The locus spans a region of 70 kb and so targeting of the hetero-specific *lox* sites necessitated two, sequential targeting events with separate constructs.

The first targeting event delivered the *lox511* site within a targeting construct containing the sequences, *lox511/hygromycin/I Sce I/lox511*, into the E14Tg2a ES cell line (HPRT⁻). The hygromycin gene was used to select for ES cells that had stably incorporated the targeting construct into the genome. The hygromycin gene was *floxed* with *lox511* sites to enable it to be deleted by Cre-mediated SSR and therefore permit its subsequent use as required following large interval RMCE. The use of the I *Sce* I restriction site will be discussed in detail later (see Chapter 3).

The second targeting event delivered the *loxP* site, as part of the *frt/I Sce I/5'hprt/loxP/neomycin* cassette. As above, the use of the I *Sce* I site will be discussed later, as will the *frt* site (see Chapters 3 and 5 respectively). The incorporation of this cassette was selected for using G418 supplemented media, that selected for the presence of the neomycin gene. The 5'hprt/*loxP* sequences derive from the hprt^{-Δ3'}/*loxP* cassette (Smith et al., 1995), but will be referred to as 5'hprt/*loxP* throughout this thesis. The 5'hprt/*loxP* cassette underwent Cre-mediated SSR with a *loxP/3'hprt (loxP/hprt^{-Δ5'})* cassette to reconstruct a functional hprt minigene, which was shown to be capable of conferring resistance to HAT selection in HPRT⁻ ES cells (Smith et al., 1995). The *loxP* site therefore had a dual role in the context of the strategy developed within this thesis: along with the *lox511* site, the *loxP* site defined the interval exchanged during RMCE and additionally it mediated the reconstruction of the hprt minigene, which was implemented as part of the selection strategy for the large interval RMCE event.

2.2 The positions of the target sites

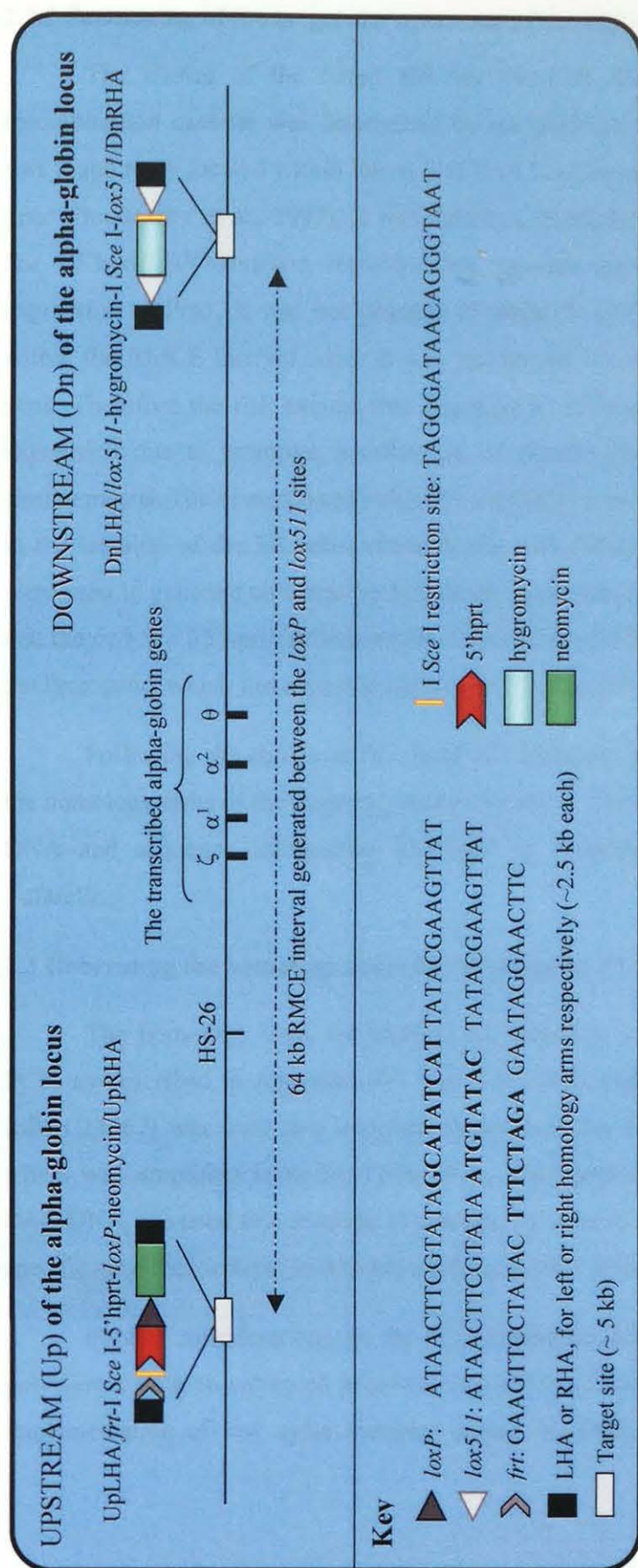
As mentioned above, two targeting events were performed, which raises the issue of nomenclature in order to specify the sites that were targeted. To address this issue, the target site for the first targeting event, which delivered *lox511/hygromycin/I Sce I/lox511*, was defined as ‘downstream’, which describes its location as downstream or 3’ to the last structural gene of the alpha-globin locus. Whilst the second targeting event, which delivered *frt/I Sce I/5’hprt/loxP/neomycin*, was defined as ‘upstream’, to describe its location as upstream or 5’ of the MRE region of the alpha-globin locus. Fig. 2.1 summarises the nomenclature, position and orientation of the targeting constructs whose construction and targeting is detailed in this chapter.

An important feature of large interval RMCE over conventional genome engineering technology is that the relevant recombination sites and selection markers can be located at positions outwith the locus of interest. This therefore reduces the potential of these sequences to interfere with the normal functioning of the locus of interest, thereby permitting a more reliable phenotypic report of any mutations purposefully made to the locus. For these reasons, careful consideration was given to choice of the location for each of the target sites.

2.2.1 Positioning the target site downstream of the alpha-globin locus

The inspection of the alpha-globin sequence data determined that the terminal functional sequence of the alpha-globin locus was the structural gene, 3’ theta (3’ θ), as depicted in Fig. 2.1. Sequence inspection also determined that the region downstream of theta was devoid of any known gene sequences or sequences known to be associated with gene regulation. It was therefore decided to target the *lox511/hygromycin/I Sce I/lox511* recombination cassette 6.9 kb downstream of theta’s terminal exon 3. It was anticipated that this position would not conflict with either the normal functioning of the alpha-globin locus nor any genes further downstream.

Fig. 2.1 Arrangement and nomenclature used for the alpha-globin locus and ES cell recombination cassettes



The arrangement of the murine alpha-globin locus on chromosome 11 is as shown (not to scale). The four genes transcribed from the cluster are represented with the symbols ζ , α and θ . Approximately 26 kb 5' of the first structural gene in the murine locus the *DNase* I hyper-sensitive site HS-26 is located, as shown. The analogous site in the human is HS-40, situated 40 kb 5' of the first structural gene of the human alpha-globin cluster, which is located on chromosome 16. HS-26 and HS-40 genetically signpost a region thought to be involved in the correct temporal expression of the structural genes of the alpha-globin cluster, known as the Major Regulatory Element (MRE), which resides within intron 5 of Prox1. The position of the targeting constructs used to deliver hetero-specific *lox* sites and the sequences necessary to select for the large interval RMCE event, are as shown. The nomenclature associated with targeting the region 3' of the alpha-globin locus was described as 'downstream' and the region 5' of the locus 'upstream', as shown. The homology arms generated to target the locus were described as either the left or right homology arm, LHA or RHA respectively. Thus DnLHA and DnRHA refer to the homology arms used to target the downstream region. Targeting was verified by Southern blot analysis with flanking probes: downstream left probe and right probe, DnLP and DnRP respectively. Similarly UpLHA or UpRHA and UpLP or UpRP denote the homology arms and probes used for the upstream targeting position.

2.2.2 Positioning of the target site upstream of the alpha-globin locus

The choice of the target site for the *frt/I Sce I/5'hprt/loxP/neomycin* recombination cassette was determined by the position of the HS-26 site. HS-26 was found to be located within intron 5 of Prox1, a known constitutively expressed gene (Bouhassira et al., 1997). It was therefore important that targeting of the *frt/I Sce I/5'hprt/loxP/neomycin* recombination cassette should not compromise the expression of Prox. It was not possible to target 5' of Prox, thereby including it within the RMCE interval, since it was juxtaposed by just 150 bp with the MPG gene. Therefore the risk existed that targeting 5' of Prox would either disrupt its expression due to promoter interference, or disrupt the expression of MPG by similar means. The disruption to either Prox or MPG could potentially have resulted in the lethality of the ES cells, either during cell culture or development, if they were used to generate chimeras by blastocyst contribution. It was therefore decided that the *frt/I Sce I/5'hprt/loxP/neomycin* construct would be targeted into intron 6 of the Prox gene, which lies approximately 1 kb 5' of the HS-26 site.

Following the choice of the target site locations it was decided to generate the homology arms of the targeting vectors by PCR. This made full use of the BAC DNA and sequence information provided by collaborators D. Higgs and C. Tufarelli.

2.3 Generating the homology arms for targeting by PCR

The homology arms for each of the targeting vectors were generated by PCR, as described in Appendix IV. The BAC harbouring the alpha-globin locus (pBAC14567) was used as a template for most of the PCRs (except for UpLHA, which was amplified from E14Tg2a DNA, see Appendix IV.e for details). The BAC DNA was used as a template if possible, in order to reduce the chance of non-specific amplification due to non-homologous primer annealing.

Further considerations on the PCR conditions included using *Pfu* turbo, a polymerase with an enhanced proof-reading ability (Lundberg et al., 1991) and the implementation of low cycle numbers during the PCR. These precautions were

taken to reduce the risk of introducing mutations to the homology arms during cycling. Mutations introduced during PCR cycling would have decreased the isogenicity between the homology arms of the targeting vector and that of the target site, which may have lead to a decrease in targeting efficiency (Deng C, 1992). In addition the introduction of mutations during PCR would have been transferred undetected to the target site and may have created an unintentional mutated phenotype.

The low cycle numbers created difficulties in generating sufficient PCR products for subsequent cloning procedures to build the targeting vectors. However, the construction of the targeting vectors was facilitated due to the implementation of non-compatible restriction sites in their construction design. The plasmid cloning vector, pBS(KS+) was restricted with non-compatible enzymes, in preparation for the insertion of a PCR product that had been similarly restricted with the equivalent enzymes. The restriction sites at the ends of each PCR product were generated by including the necessary sequences for their creation at the 5' end of each primer pair. This approach reduced the number of background colonies that were generated from the ligation between the plasmid and the PCR products and allowed the orientation of cloned PCR product to be pre-determined.

The homology arms for each targeting construct were first built into a plasmid before the recombination cassettes, which were constructed in parallel, were introduced, (see Figs. 2.2 and 2.6). The PCR strategy was designed in each case to construct homology arms that were flanked by a *Not* I and a *Sal* I site. At the centre of the each homology arm pair was a *Bam*HI site, which was created from sequences originating from the primers used for their generation. This site was the position into which the relevant recombination cassette was introduced. Either the *Not* I or *Sal* I site was used to linearise the completed targeting vector, prior to its electroporation into ES cells.

The use of *Bam*HI, *Sal* I and *Not* I restriction sites to clone the targeting constructs therefore placed a constraint on the target site that was selected, since it was required to be devoid of these sequences. Therefore for both target sites a

genomic region, of approximately 5 kb, that lacked *Bam*HI, *Not* I and *Sal* I sites was selected.

2.4 Targeting the *lox511*/hygromycin/I *Sce* I/*lox511* construct downstream of the alpha-globin locus

The aim of the first targeting event was to deliver *lox511*/hygromycin/I *Sce* I/*lox511* to the target site downstream of the alpha-globin locus. The construction of the targeting vector is illustrated in Fig. 2.2. At the point of electroporating the targeting vector containing *lox511*/hygromycin/I *Sce* I/*lox511*, the orientation the *loxP* site, that would be subsequently delivered upstream of HS-26, was unconfirmed. This is because the construction of the *frt*/I *Sce* I/5'hpvt/*loxP*/neomycin construct was incomplete at that time. Therefore to retain flexibility for subsequent targeting events both orientations of the *lox511*/hygromycin/I *Sce* I/*lox511* constructs, were electroporated, they are each illustrated as CP# 47 and CP# 48 in Fig. 2.2.

5×10^7 E14Tg2a ES cells were electroporated with 75 μ g of *Not* I linearised targeting vector for each of CP# 47 and CP# 48. Following electroporation cells were plated at densities of 2.5×10^6 , 1.25×10^6 and 6.25×10^5 per 10cm petri dish. 48 hours after electroporation hygromycin selection was added to the culture media.

Colonies that appeared after 12 days in hygromycin selection were picked into 96 well plates. Three 96 well plates were picked for each of the targeting constructs. Plates were cultured to adequate density and then split 1:2, one plate was frozen and one was used to prepare DNA. DNA was prepared from the 96 well clones and screened for targeting by Southern blot analysis. Fig. 2.3 illustrates the Southern blotting screening strategy for the *lox511*/hygromycin/I *Sce* I/*lox511* construct and Fig. 2.4 shows the results of the analysis for CP#48.

Following the preliminary screening of 96 well DNA a few clones were selected for further Southern blot analysis to confirm that the desired targeting event had occurred, see Fig. 2.5. The targeting efficiency for CP# 47 was 18% and for CP# 48 was 15%, of hygromycin^R clones that were determined as targeted by

Fig. 2.2 Construction of the ES cell targeting vectors CP#47 and CP#48, which contain the sequences for *lox511*/hygromycin/*I Sce I*/*lox511*

Oligos Z6205 and Z6207 were annealed to form a double-stranded DNA fragment that contained two *lox511* sites (in the same orientation as each other) and 5' single-stranded overhanging ends that permitted its ligation into a *Bam*HI restricted site. The annealed oligos were ligated into *Bam*HI restricted pBS(KS+) and the resulting colonies were screened by colony hybridisation, with 5' ³²P labelled Z2605 as the probe. The orientation of the *lox511* sites was determined by sequencing with T3 and T7 primers (see Fig. 2.3a). The resultant plasmids with alternative orientations of the inserted oligos were termed CP#20 and CP#21. A *Bgl*II fragment from pHA58 (AJHS#948) was generated, which included the hygromycin resistance gene, driven by the murine phosphoglycerate kinase (PGK) promoter and the necessary polyadenylation addition sequences. This *Bgl*II fragment was ligated into CP#20, which had been previously restricted with at a unique *Bgl*II site, created within the Z6205 and Z6207 oligos. The orientation of hygromycin gene was determined by sequencing with T3 and T7 primers and the resulting plasmid termed CP#41a (see Fig. 2.2b). Oligos A4128 and A4129, which contained the sequence of the *I Sce I* restriction site, were annealed and ligated into *Bgl*II partially restricted CP#41a. Introduction of the *I Sce I* site was screened by colony hybridisation, with 5' ³²P labelled A4128 as the probe. The positive plasmids were pooled and termed CP#41b (see Fig. 2.2c), this pool represented a mixture of plasmids with the possibility of harbouring the *I Sce I* site at either side of the hygromycin gene, however the precise location of the *I Sce I* site in the final targeting vector was determined by sequencing (see below). The homology arms for this targeting vector were built into a separate plasmid. DnLHA was the PCR product from primers D6624 and D6625 (see Appendix IV.g) and was restricted with *Sal*I and *Bam*HI and then ligated into a similarly restricted pBS(KS+) plasmid to generate CP#44 (see Fig. 2.2d). Next, CP#44 was *Bam*HI and *Not*I restricted and the PCR product from primers D6626 and D6627, which generated DnRHA (see Appendix IV.h), was similarly restricted and cloned into this site to give CP#45 (see Fig. 2.2e). CP#45 was then *Bam*HI restricted at the unique site created between the DnLHA and DnRHA and the 1.6 kb *Bam*HI fragment from CP#41b, containing the *lox511*/hygromycin/*I Sce I*/*lox511* sequences, ligated into this site. The orientation of the inserted *Bam*HI fragment was determined by *Sph*I and *Stu*I restriction and the orientation of the *lox511* sites and the location and of the *I Sce I* site was confirmed by sequencing with primers F1346 and F1347. The two alternative orientations of the *lox511* sites with respect to the homology arms were obtained and termed, CP#47 or 'HYG-3' (see Fig. 2.2f) and CP#48 or 'HYG-4' (see Fig. 2.2g).

Fig. 2.2 Construction of the ES cell targeting vectors CP#47 and CP#48, which contain *lox511*/hygromycin/*I Sce I*/*lox511*

Fig. 2.2a

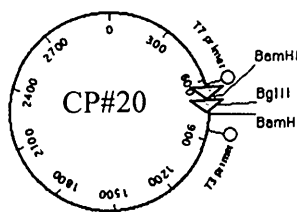


Fig. 2.2b

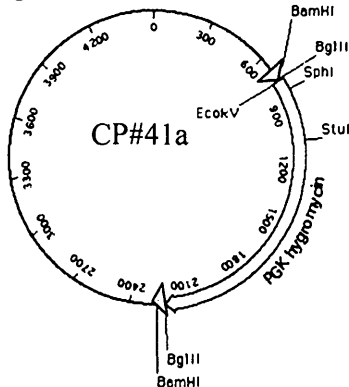


Fig. 2.2c

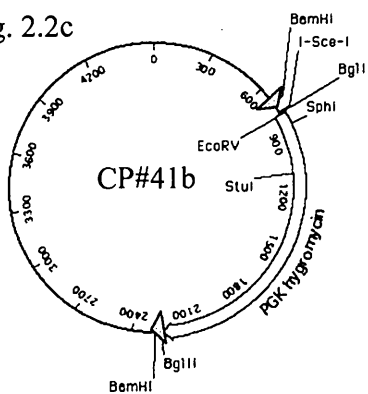


Fig. 2.2d

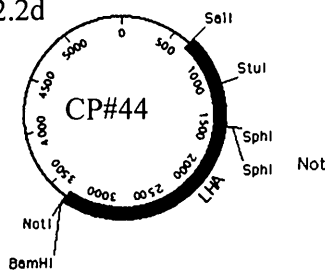


Fig. 2.2e

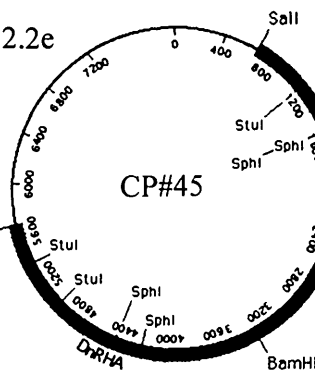


Fig. 2.2f

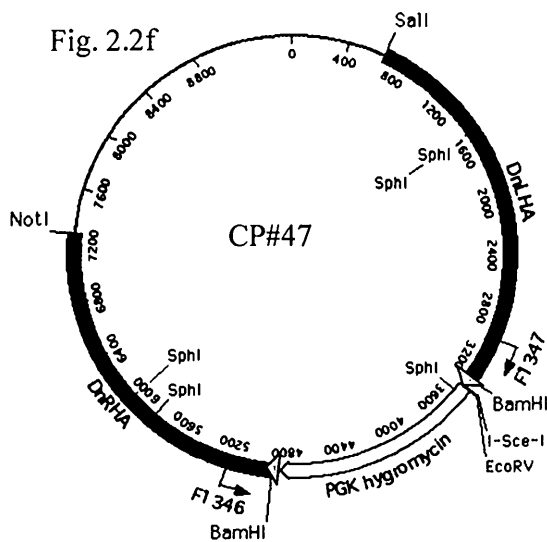


Fig. 2.2g

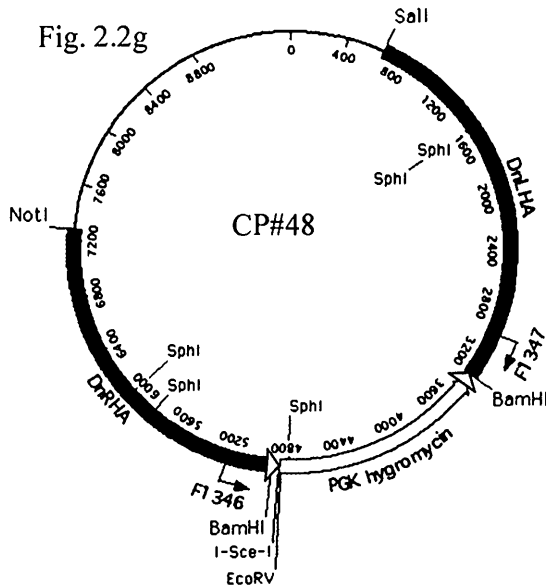


Fig. 2.3 Southern blot screening strategy employed to detect targeting of *lox511/hygromycin/I Sce I/lox511*

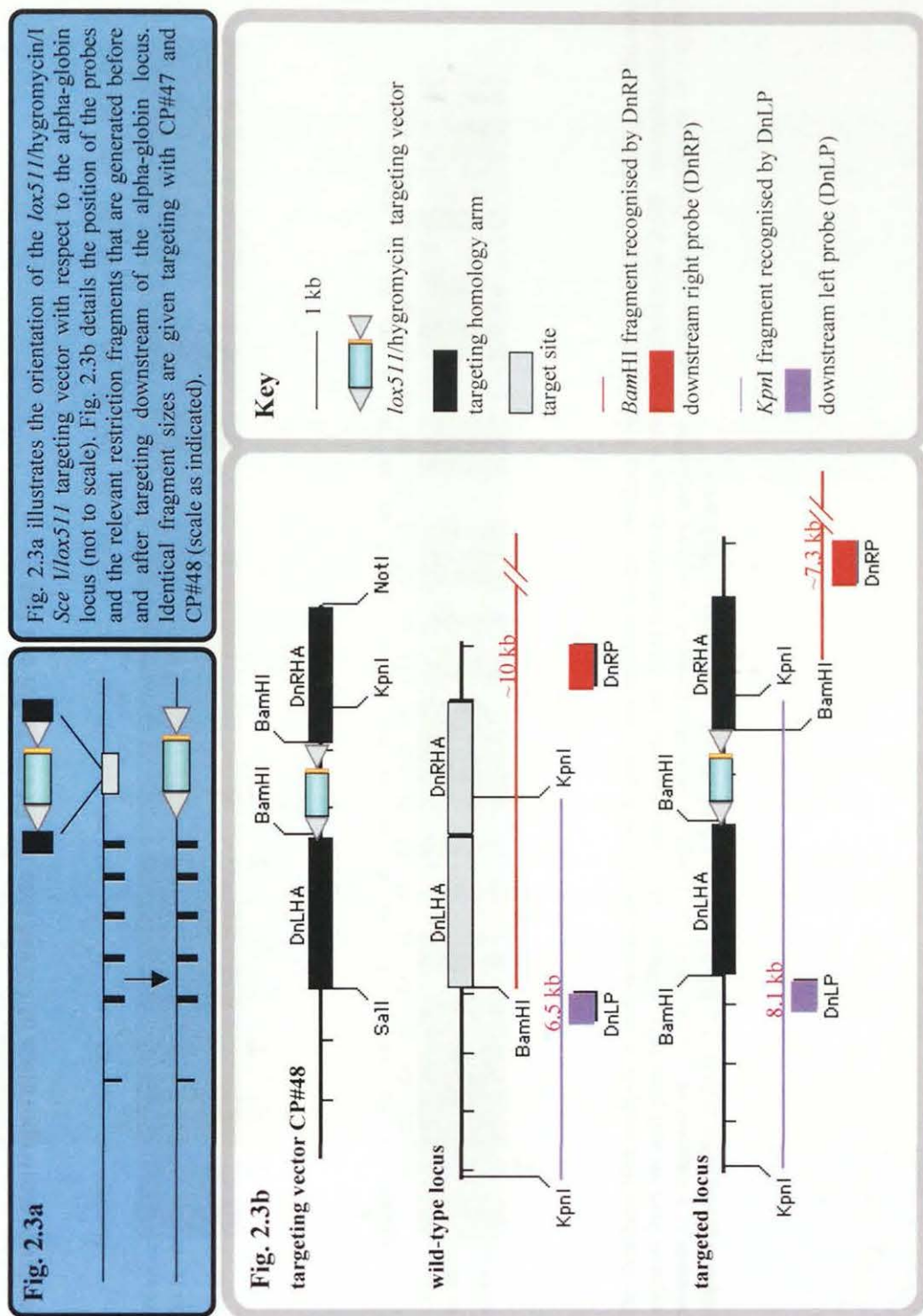
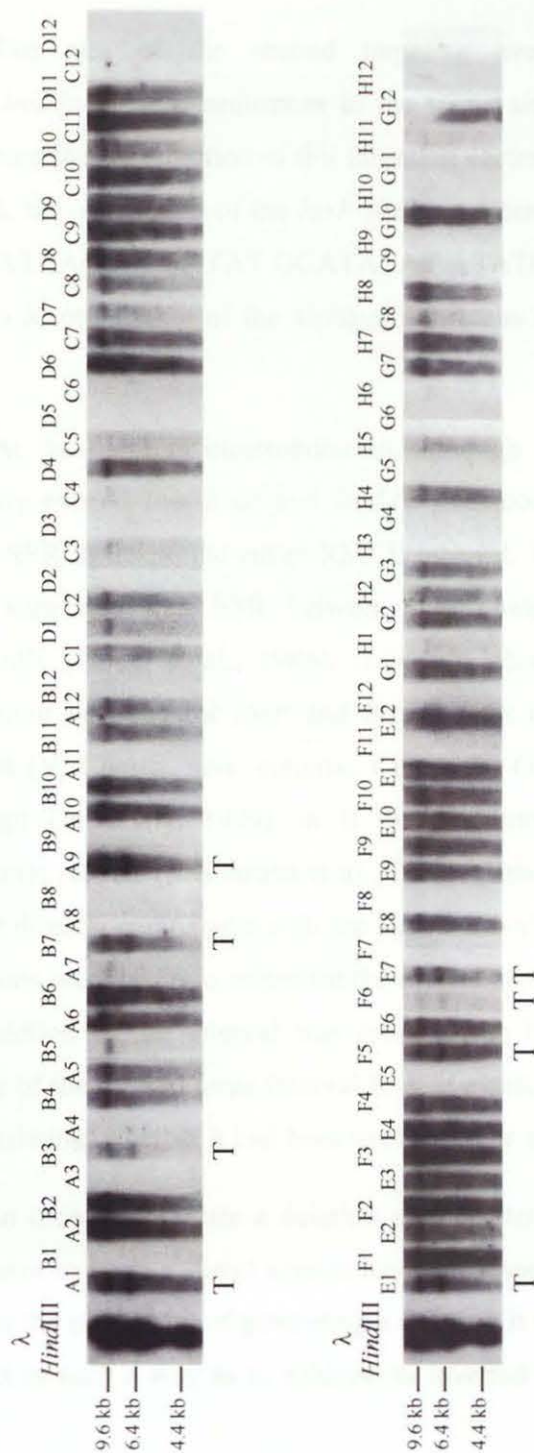


Fig. 2.4 A southern blot screen of *Bam*HI restricted genomic DNA prepared from a 96 well plate of hygromycin resistant ES cell colonies that arose from electroporation of CP#48 into the E14Tg2a ES cell line



The Southern blot analysis shown was performed on hygromycin resistant clones resulting from the electroporation of E14Tg2a cells with CP#48. DNA was prepared from 96 well plate HYG-4/Plate 2. DNA was digested with *Bam*HI and the Southern blot hybridised with DnRP. An untargeted locus is indicated by the presence of a fragment of ~10 kb, whilst for a targeted locus this fragment is reduced in size by 2.7 kb, generating a fragment of ~7.3 kb. Clones that harbour a single targeting event include: A1, B3, B7, A9, E1, F5, F6 and E7, as indicated by the letter T.

Southern blotting. The Southern blot data indicate that clones A1, A9, E7 and F5 from CP#48 (HYG-4/Plate2) harboured the desired targeting event.

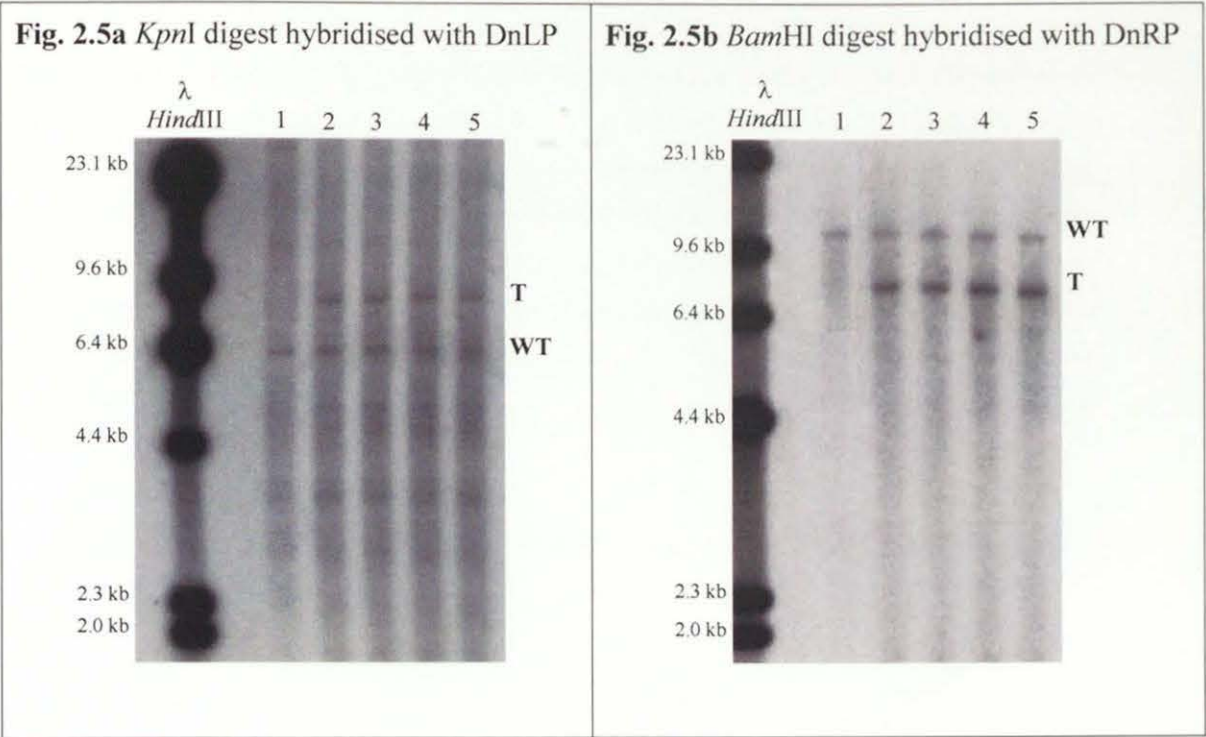
2.5 Targeting the *frit/I Sce I/5'hpvt/loxP*/neomycin construct upstream of the alpha-globin locus

The aim of the second targeting event was to deliver *frit/I Sce I/5'hpvt/loxP*/neomycin sequences to the target site upstream of the alpha-globin locus. Once the construction of this targeting vector was completed (as described in Fig. 2.6), the orientation of the *loxP* site was determined by sequencing and found to be 5'ATAACTTCGATAT GCATACAT ATATCGAAGTTAT3' with respect to the 5' to 3' orientation of the alpha-globin locus sequences, as illustrated in Fig. 2.1.

At the point of electroporation, although considered highly unlikely, the possibility existed that *loxP* and *lox511* sites could undergo Cre-mediated non-specific SSR to delete the entire RMCE interval. The earliest studies on *loxP* and *lox511* suggested that SSR between these hetero-specific *lox* sites occurred infrequently (Hoess et al., 1986). However other evidence suggested that SSR might occur between the *loxP* and *lox511* sites if the level of Cre activity was sufficient (Y. Cherifi, pers. comms., Centre for Genome Research Staff Seminar, Edinburgh University, 1999), or if the environment of the *lox* site contained palindromic repeats (Mlynarova et al., 2002). Although these reported experiments were not directly comparable with the large interval RMCE developed in this thesis, precautions were taken to minimize the effects of non-specific SSR. This was done since deletion of the interval was predicted to be a significant obstacle to the recovery of the desired large interval RMCE event, as it would prevent the interval from persisting, whether it had been exchanged or not.

In order to mediate a deletion of the interval the hetero-specific *lox* sites would have to be in a direct orientation with respect to one another. Therefore, to eliminate the possibility of generating a deletion it was decided to target the second construct in such a way as to achieve an inverted orientation of the *lox* sites. The

Fig. 2.5 Southern blotting on putatively targeted clones obtained from electroporation of E14Tg2a with CP#48 (from 96 well plate HYG-4/Plate 2)

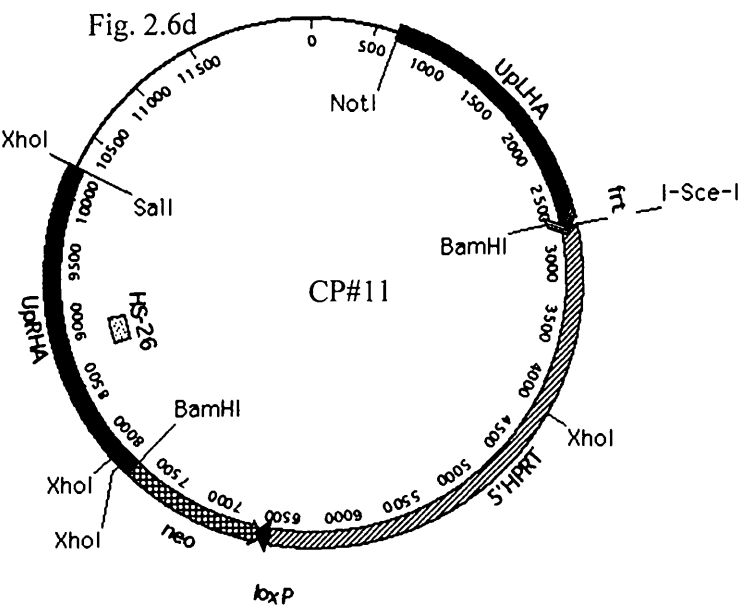
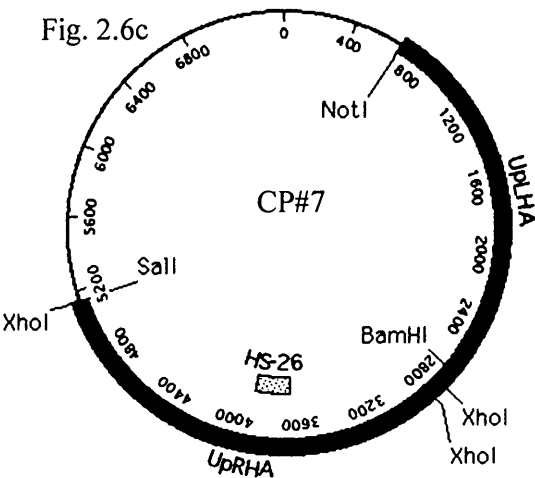
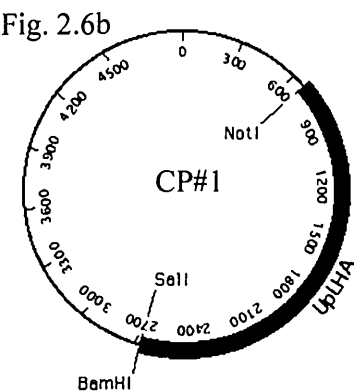
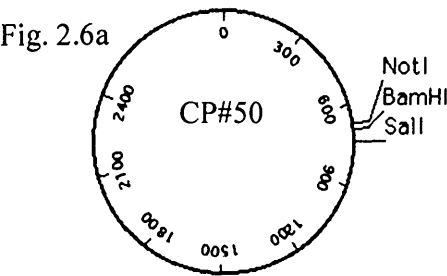


Lane 1, E14Tg2a; Lane 2, clone A1; Lane 3, clone A9; Lane 4, clone E7; Lane 5, clone F5.

**Fig. 2.6 Construction of the ES cell targeting vector CP#11, which contains *frt*/I
Sce I/5'hppt/neomycin**

The upstream left homology arm (UpLHA) and upstream right homology arm (UpRHA) were generated by PCR, as outlined in IV.e and IV.f respectively. The PCR generated UpLHA DNA fragment was restricted with *Bam*HI and *Not*I and then cloned in to pBS(KS+) which had been similarly restricted to yield CP#1 (see Fig. 2.6b). Next, the PCR generated UpRHA DNA fragment was restricted with *Bam*HI and *Sal*I and cloned into similarly restricted CP#1, to yield CP#7 (see Fig. 2.6c). CP#7 was restricted with *Bam*HI and treated with alkaline phosphatase before the *Bam*HI fragment from AJHS#1171, as shown, was cloned into this site. The desired orientation of the AJHS#1171 derived fragment was determined by *Xho*I restriction and sequencing was performed to verify the orientation and integrity of the *loxP* site. The resultant plasmid was termed CP#11 (see Fig. 2.6d) and was linearised by *Sal*I restriction prior to electroporation into ES cells.

**Fig. 2.6 Construction of the ES cell targeting vector CP#11, which contains *frt*/I
Sce I/5'hppt/neomycin**



inverted orientation should produce an inversion of the locus rather than a deletion, should non-specific SSR occur. It was predicted that the inversion reaction would be reversible, with equal probability, the consequence of which would be that roughly 50 % of the clones would harbour an inverted locus and 50 % should be normal. In order to distinguish these alternative genomic arrangements Southern blotting could be used.

As described previously, the first targeting event generated lines that harboured both orientations of the *lox511*/hygromycin/*I* *Sce* *I*/*lox511* constructs. To produce a cell line with the required inverted *lox* orientation when targeted with *frt*/*I* *Sce* *I*/5'hppt/*loxP*/neomycin, a clone targeted with the CP#48 targeting vector was therefore required.

The targeting vector containing the *frt*/*I* *Sce* *I*/5'hppt/*loxP*/neomycin targeting construct was targeted into Clone A9 from HYG-4/Plate 2 to generate a double-targeted ES cell line in which the *lox* sites were in an inverted orientation with respect to one another.

The homology arms used to target the *frt*/*I* *Sce* *I*/5'hppt/*loxP*/neomycin recombination cassette were generated by PCR. The alpha-globin sequence residing in pBAC14567 terminated just 199 bp upstream of the desired target site for the *frt*/*I* *Sce* *I*/5'hppt/*loxP*/neomycin recombination cassette. Therefore there was a lack of alpha-globin sequence information from pBAC14567 in order to design one of the primers necessary to PCR generate a 2.5 kb fragment for the left homology arm (UpLHA). To overcome the limitation of sequence availability the project collaborator, D. Higgs, provided the upstream sequence of a contig cosmid to pBAC14567. Additionally the cosmid DNA was used as a template to generate the probe, UpLP (see Appendix IV.a), which was used to screen the Southern blots for this targeting vector.

The ES cell line HYG-4/Plate2/Clone A9 was electroporated with 50 µg of *Sal* I linearised targeting DNA. 5×10^7 cells were electroporated and then plated at densities of 2.5×10^6 /10cm dish and 1.25×10^6 /10cm dish, in culture media with

G418 selection added. After 10 days of selection G418^R colonies were picked into 96 well plates.

The 96 well plates were cultured and split 1:2, one plate was frozen and one used to prepare DNA for Southern blot analysis. The screening strategy for the *frt/I Sce I/5'hprt/loxP/neomycin* construct is shown in Fig. 2.7, whilst the 96 well screens are shown Fig. 2.8a and Fig. 2.8b. The targeting efficiency for CP# 11 into HYG4-A9 was 22%, of G418^R clones that were shown to have targeted by Southern blotting. Clones suspected of being double-targeted were termed "Completed Alpha-globin Targeted (CAT)- clone number", e.g. CAT-A3. Four CAT clones from the 96 well DNA screen were selected for further Southern blot analysis to confirm that the desired targeting event had occurred, see Fig. 2.9.

2.6 Summary of ES cell targeting

This chapter has described the generation of double-targeted 'CAT' ES cell line, by the sequential targeting of *lox511/hygromycin/I Sce I/lox511*, followed by *frt/I Sce I/5'hprt/loxP/neomycin* into an E14Tg2a ES cell line. The generation of these ES cell lines was part of the development of a strategy of large interval RMCE. The aim of targeting was to deliver hetero-specific *lox* sites, in an inverted orientation, into the ES cell genome at positions that would flank a single-copy of the alpha-globin locus, thereby creating an interval of approximately 64 kb.

However, due to the diploid nature of E14Tg2a cells there exist two copies of the alpha-globin locus, each residing on chromosomes 11. Therefore providing each targeting event occurred independently of each other, it was anticipated that 50% of the CAT clones would have targeted both constructs to the same chromosome and 50% would have targeted the constructs to alternate chromosomes. The large interval RMCE strategy depends upon both constructs being targeted to the same chromosome, in order to generate an alpha-globin locus that is flanked by both hetero-specific *lox* sites. The following chapter describes the further characterisation of the CAT clones in order to determine their targeting arrangement in this respect.

Fig. 2.7 Southern blot screening strategy employed to detect targeting of *frr1* *Sce* I/5'hprrt/*loxP*/neomycin

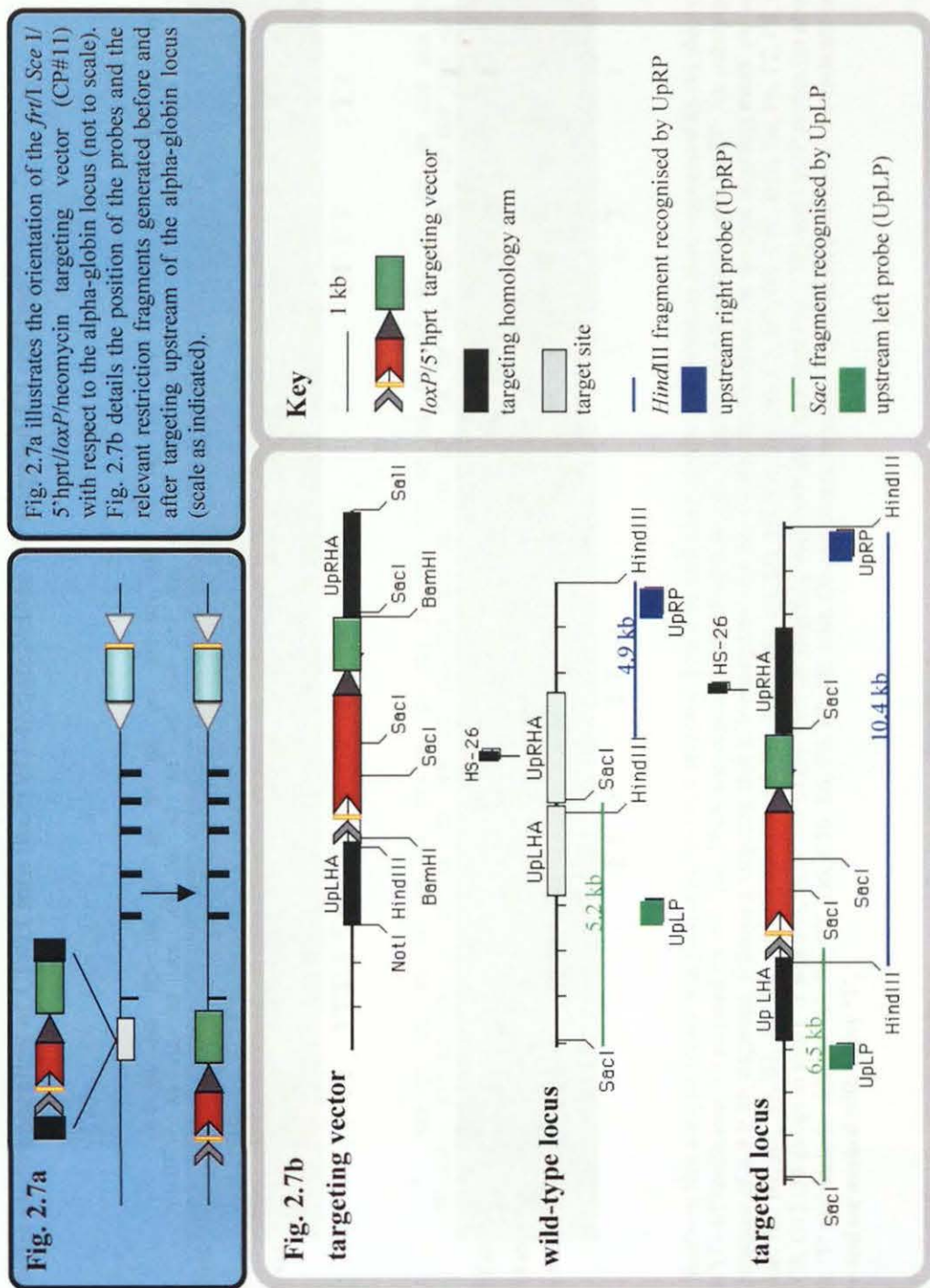
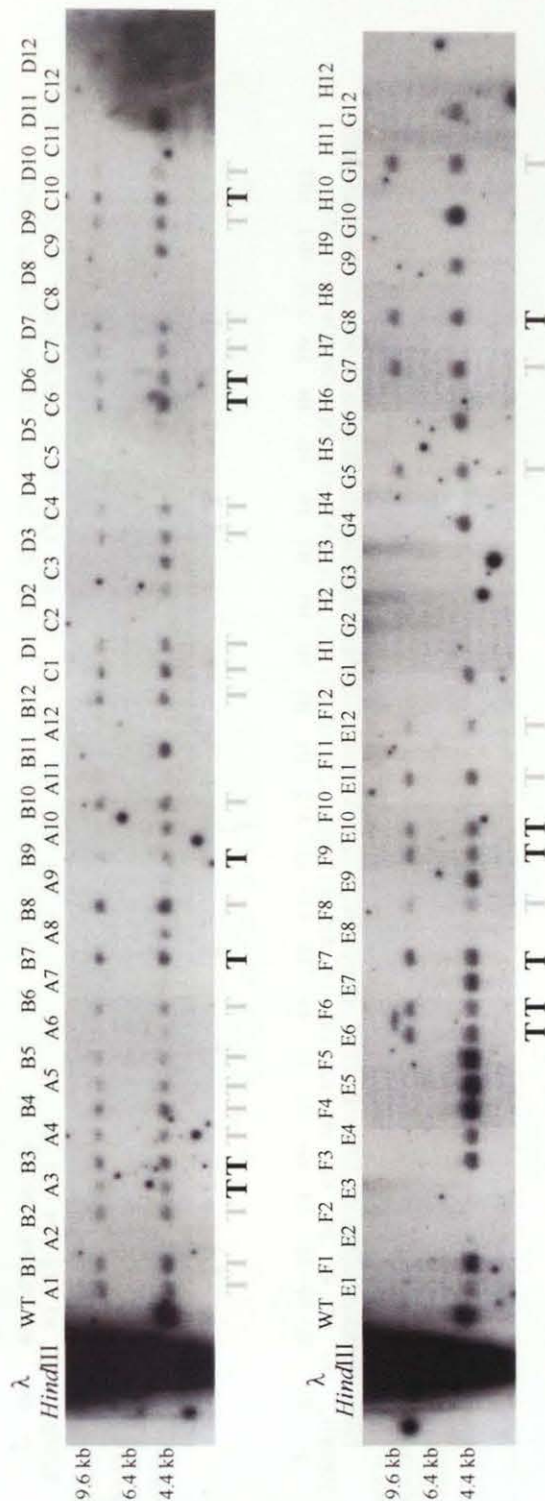
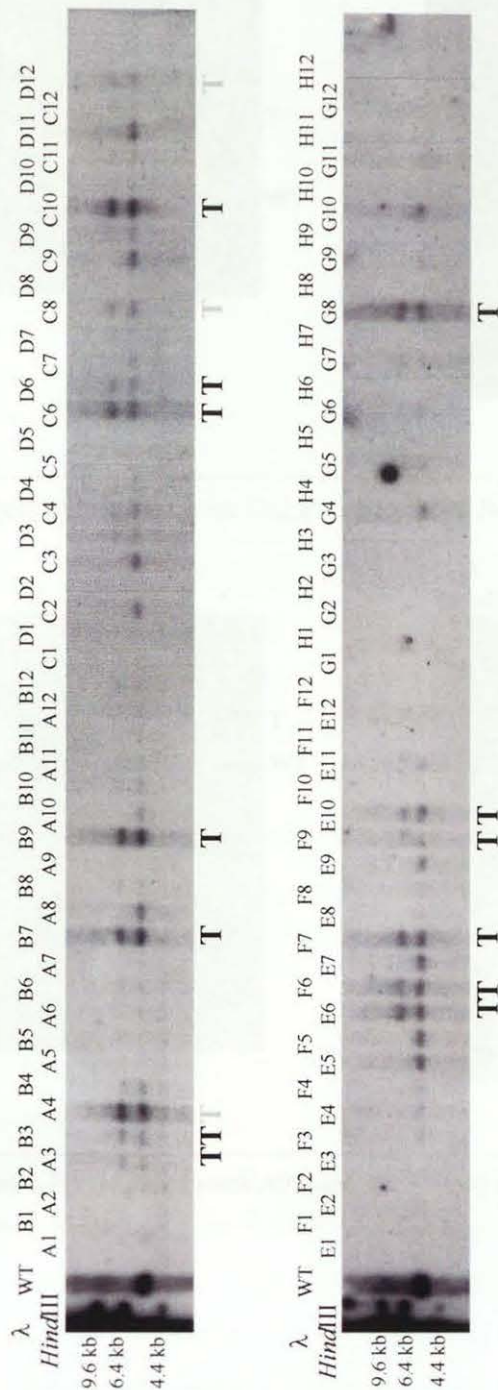


Fig. 2.8a A southern blot screen of *Hind*III restricted genomic DNA prepared from a 96 well plate of neomycin resistant ES cell colonies that arose from electroporation of CP#11 into the HYG-4/Plate2/clone A9 targeted ES cell line



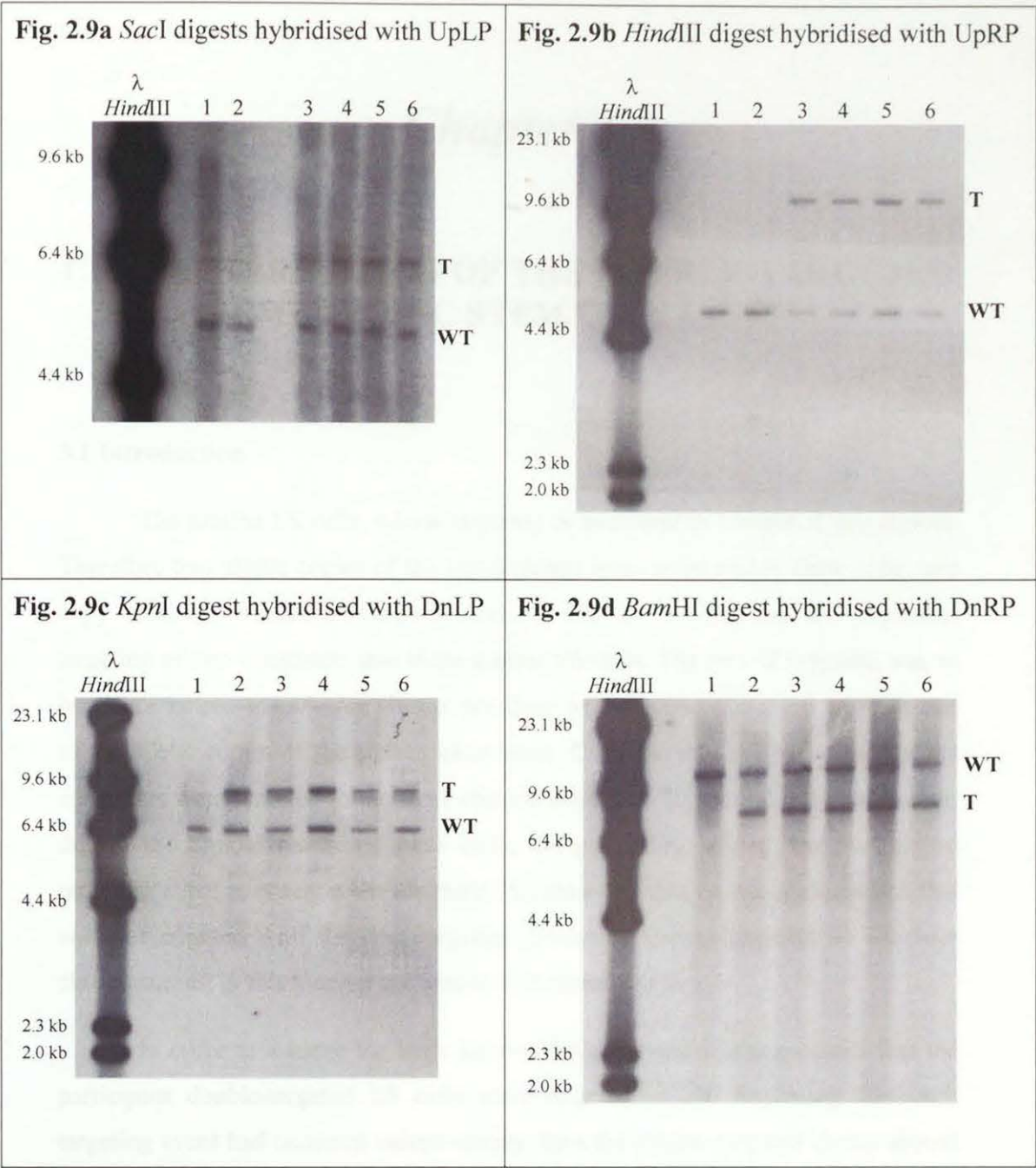
The Southern blot analysis shown was performed on DNA prepared from a 96 well plate of neomycin resistant clones generated by the electroporation of CP#11 into the HYG-4/Plate2/clone A9 targeted ES cell line. DNA was digested with *Hind*III and the Southern blot hybridised with UpRP. An untarged locus is indicated by the presence of a 4.9 kb fragment, following targeting this is increased to 10.4 kb. Clones that indicate the desired targeting event are indicated with a 'T' and include: A1, B1, B2, A3, B3, A4, B4, A5, B5, B6, B7, B8, B9, B10, B12, C1, D1, D3, C4, C6, C7, D7, D9, C10, D10, E6, F6, F7, F8, F9, E10, E11, E12, G5, G7, G8, G11 (38 clones in total). In addition, those clones that indicate targeting with both the UpLP (see Fig. 2.8b) and UpRP probes (as above) are indicated with a black 'T' and include: A3, B3, B7, B9, C6, D6, C10, E6, F6, F7, F9, E10, G8 (13 clones in total). The remainder of the clones indicate targeting with the UpRP alone and are marked with a grey 'T'.

Fig. 2.8b A southern blot screen of *SacI* restricted genomic DNA prepared from a 96 well plate of neomycin resistant ES cell colonies that arose from electroporation of CP#11 into the HYG-4/Plate2/clone A9 targeted ES cell line



The Southern blot analysis shown was performed on DNA prepared from a 96 well plate of neomycin resistant clones generated by the electroporation of CP#11 into the HYG-4/Plate2/clone A9 targeted ES cell line. DNA was digested with *SacI* and the Southern blot probed with UpLP. An untargated locus is indicated by the presence of 5.2 kb fragment, following targeting this is increased to 6.5 kb. Clones that indicate the desired targeting event are indicated with a 'T' and include: A3, B3, A4, B7, B9, C6, D6, C8, C10, D12, E6, F6, F7, F9, E10, G8 (16 clones in total). In addition, those clones that indicate targeting with both the UpRP (see Fig. 2.8a) and UpLP probes (as above) are indicated with a black 'T' and include: A3, B3, B7, B9, C6, D6, C10, E6, F6, F7, F9, E10, G8 (13 double-targeted clones in total). The remaining clones indicate targeting with the UpLP alone, are indicated with a grey 'T'.

Fig. 2.9 Southern blotting on selected double targeted CAT ES cell lines



Lane 1, E14Tg2a; Lane 2, HYG4-plate2 clone A9; Lane 3, CAT-A3; Lane 4, CAT-B3; Lane 5, CAT-C6; Lane 6, CAT-C10.

Chapter 3

CHARACTERISATION OF THE DOUBLE-TARGETED EMBRYONIC STEM CELL LINES

3.1 Introduction

The murine ES cells, whose targeting is described in Chapter 2, are diploid. Therefore two allelic copies of the alpha-globin locus exist within these cells, one copy residing on each of chromosomes 11. Chapter 2 described the sequential targeting of two constructs into these murine ES cells. The aim of targeting was to introduce hetero-specific *lox* sites at positions in the genome in order to flank one of the allelic copies of the alpha-globin locus. Consequently, this required that the constructs were targeted to the same chromosome, that is targeted *in cis*. However, due to the diploid nature of these cells, the possibility existed that the second targeting event may target the alternate chromosome, that is, the chromosome that was not targeted with the first targeting construct. Clones targeted on alternate chromosomes, in this manner are termed as targeted *in trans*.

In order to achieve the large interval RMCE event it was essential that the participant double-targeted ES cells were targeted *in cis*. Assuming that each targeting event had occurred independently, then the double-targeted clones should represent a population of clones that harboured the *in cis* and the *in trans* targeting arrangements in equal proportions. It was therefore apparent that the double-targeted clones would require analysis to distinguish those that harboured the desired *in cis* arrangement of the targeting constructs from those that did not. The Southern blot analysis described in Chapter 2 does not distinguish the nature of

targeting in this respect and so this chapter describes the analysis that was performed on a selected number of double-targeted CAT ES cell clones in order to assign *in cis* targeting.

This chapter also describes the functional analysis of the *lox* sites included within the targeting constructs. The results of these analyses were anticipated to further verify the assignment of *in cis* targeting.

Finally, the results of the analyses described in this chapter informed the selection of the cell lines chosen to participate in the large interval RMCE event, as described in Chapter 5.

3.2 The strategy used to identify *in cis* double-targeted ES cell lines

The size of the interval created between constructs targeted *in cis*, as desired, was determined by sequence inspection to be approximately 64 kb. The generation of this 64 kb interval as a restriction fragment and its subsequent detection by Southern blot analysis formed the basis of the strategy implemented here to identify *in cis* targeted clones. To affect this strategy, the restriction sites necessary to generate the interval fragment were included in the design and construction of the targeting constructs.

Crucial to the success of this strategy was the generation of the entire interval as a restriction fragment, which would only be possible if the necessary restriction sites were present within the targeting constructs alone and absent from the interval. The occurrence of restriction sites within the interval would flaw this strategy since they would prevent the generation of the 64 kb diagnostic fragment.

The requirement to select an enzyme that did not restrict within the interval excluded the majority of those with recognition sequences of 6 to 8 bp, since they were present within the interval, as determined by sequence inspection. The variety of restriction endonucleases that could be used was therefore significantly limited.

Ultimately, to ensure that restriction occurred within the targeting constructs alone, here and for the future application of this strategy to other genomic locations,

the rare-cutting restriction endonuclease I *Sce* I was chosen. The restriction endonuclease I *Sce* I is also known as Omega-nuclease (Colleaux et al., 1988; Thierry et al., 1991). This enzyme has an 18 bp recognition sequence and so the probability of its occurrence at random is once per 6.9×10^{10} bp or once per 20 human genomes (Monteilhet et al., 1990). This ensures that the probability of its occurrence within a genome, outwith its targeted delivery, to be extremely unlikely. Therefore the likelihood of an I *Sce* I recognition site within the interval generated by *in cis* targeting was considered to be highly unlikely. Further, inspection of the sequence provided no evidence for the occurrence of this site within the interval.

In order to detect a restriction fragment of 64 kb it was necessary to separate the restricted genomic DNA by pulsed field gel electrophoresis (PFGE).

3.2.1 Southern blot analysis of restricted CAT genomic DNA separated by PFGE

Four of the double-targeted CAT lines were selected at random for analysis. Assuming that the sequential targeting events had occurred independently of each other then this sample size ensured that the probability of assigning at least one of the lines as targeted *in cis* would be above 90 %. The lines chosen were CAT-A3, CAT-B3, CAT-C6 and CAT-C10. Genomic DNA from each line was prepared for restriction and subsequent PFGE separation according to standard methods (see Appendix I).

The restriction strategy implemented here to analyse the CAT clones was designed to provide evidence for either *in cis* or *in trans* targeting. It was apparent that restriction of genomic DNA from the CAT clones with I *Sce* I alone would only provide a diagnostic restriction fragment to indicate *in cis* targeting and not *in trans*. Therefore in addition to restricting with I *Sce* I another enzyme was required that would generate restriction fragments to indicate *in trans* targeting.

The inspection of the alpha-globin sequence from the pBAC14567 sequence data determined that *Not* I did not restrict within the interval and so this enzyme was used, either alone or in combination with I *Sce* I, to generate restriction fragments that could be detected to indicate whether *in cis* or *in trans* targeting had

occurred. The *Not* I and I *Sce* I restriction sites and predicted restriction fragments are shown in Fig. 3.1.

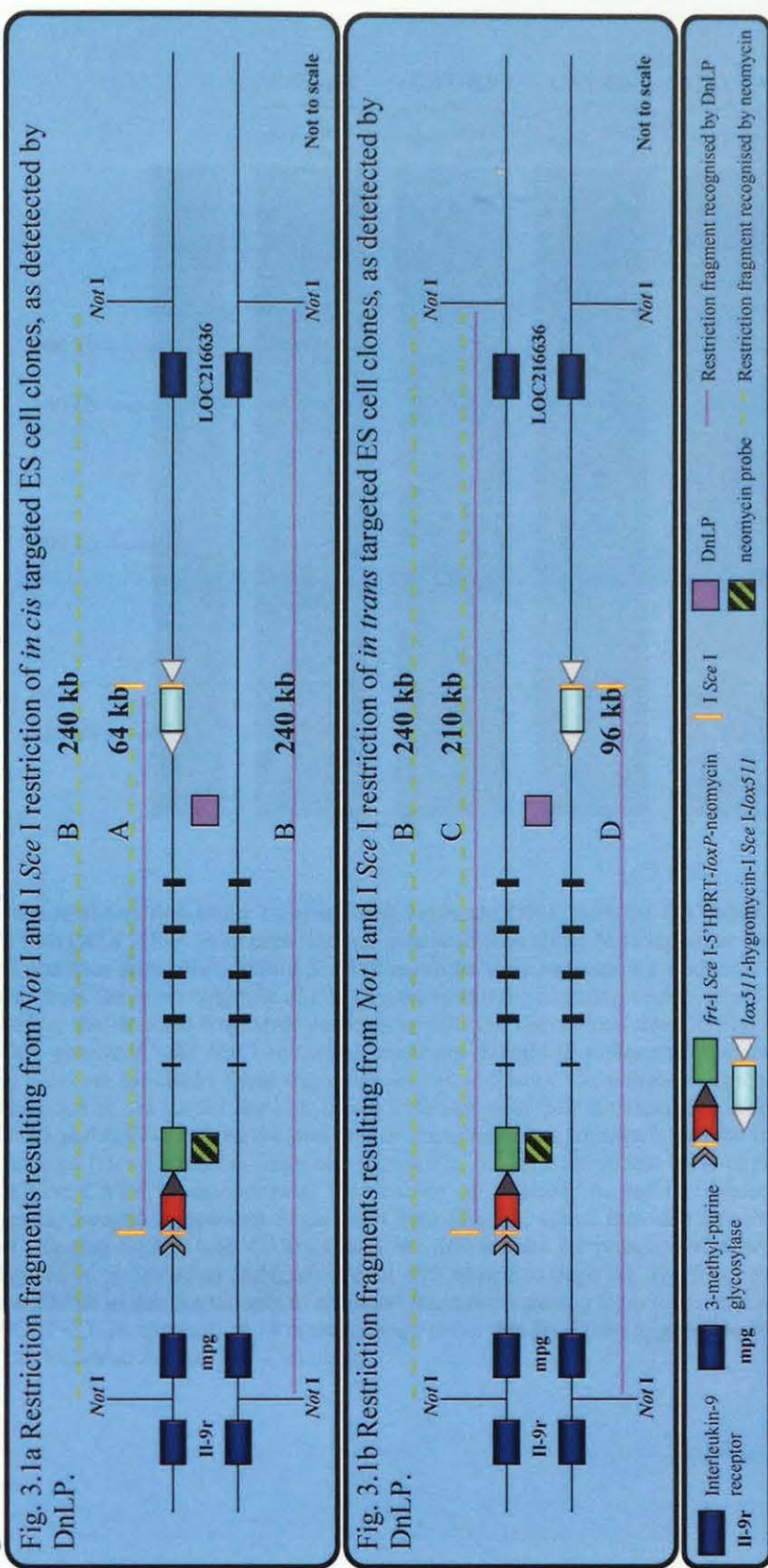
The restricted DNA was separated by PFGE, blotted and then hybridised, first with the DnLP probe (see Appendix IV), which is homologous to sequences downstream of the alpha-globin locus and internal to the RMCE interval (the results are shown in Fig. 3.2). The filter was then stripped and subsequently hybridised with a neomycin probe (supplied by Stephen Meek), which is homologous to the neomycin sequence within the *frt*/I *Sce* I/5'hprt/*loxP*/neomycin upstream targeting vector (the results are shown in Fig. 3.3). The location of both probes and the fragments they detect is illustrated in Fig. 3.1.

3.2.2 Interpretation of the PFGE results

Fig. 3.1 illustrates how the interval restriction fragment of 64 kb (fragment A) can be detected with both the DnLP and the neomycin probe, for clones that have targeted *in cis* and undergone I *Sce* I restriction. This band is clearly visible in Fig. 3.2 and Fig. 3.3 for clones CAT-A3 and CAT-B3, restricted with a combination of *Not* I and I *Sce* I. This would therefore provide positive evidence to suggest that CAT-A3 and CAT-B3 have been targeted *in cis*.

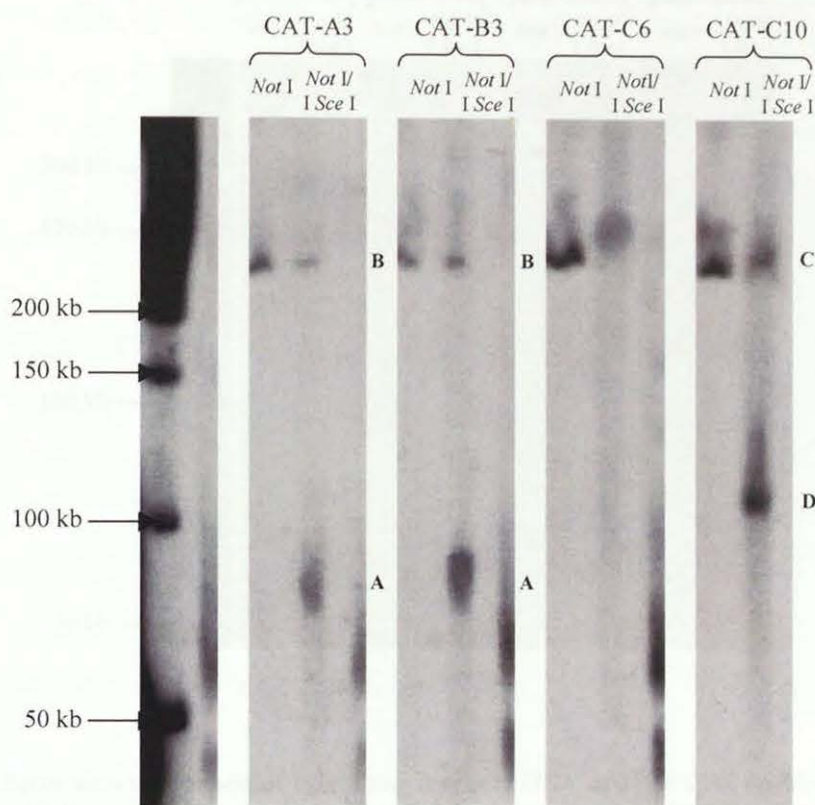
The presence of fragments, other than fragment A, for clones CAT-A3 and CAT-B3 in Fig. 3.2 and Fig. 3.3, were interpreted as additional evidence to support their assignment as *in cis* targeted clones. For instance, clones CAT-A3 and CAT-B3 shared equivalent banding patterns with respect to the fragments that appear to be greater than 200 kb in size. These fragments are thought to represent fragment B (see Fig. 3.1). Fig. 3.2 shows the detection of fragment B with the DnLP when the DNA is restricted with either *Not* I alone, or in combination with I *Sce* I. In contrast, Fig. 3.3 shows the detection of fragment B, by the neomycin probe, only when DNA is restricted with *Not* I alone. As shown in Fig. 3.3, fragment B is not detected when a combination of *Not* I and I *Sce* I restriction is performed.

Fig. 3.1 PFGE screen to determine *in cis* or *in trans* targeting of the targeting constructs



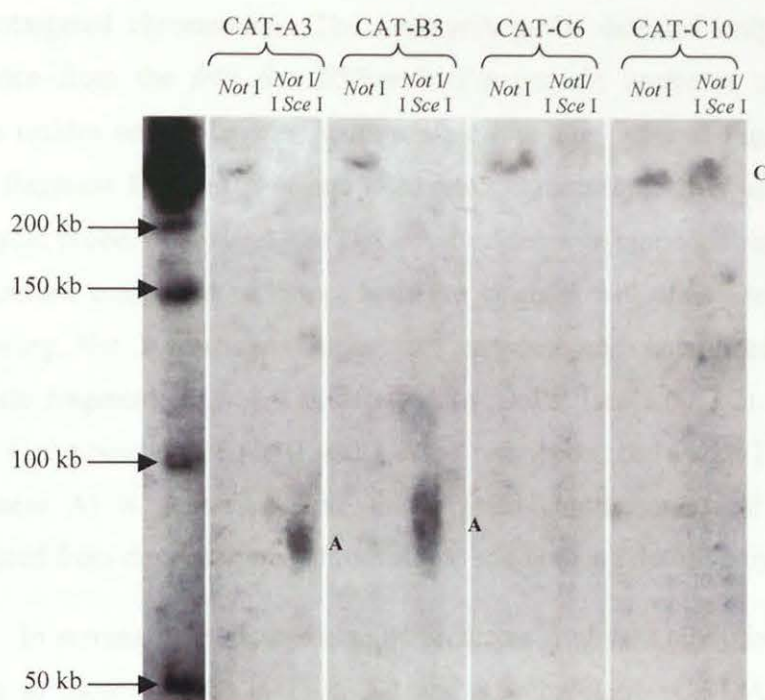
The diagram illustrates the Southern blot probing strategy used to assess whether the recombination cassettes in double-targeted ES cell clones were targeted *in cis* or *in trans*. As depicted, *Not* I sites are located outwith the RMCE interval as determined by sequence inspection of mouse chromosome 11, A4:A5, 32160K to 32450K (<http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/mapsdb.cgi?org=mouse&chr=11>). In addition, the location of the DnLP and neomycin probe are shown along with the corresponding restriction fragments that they are predicted to detect, following a combination of I *Sce* I and *Not* I restriction. Upon I *Sce* I restriction, DnLP detects fragments A and B, indicating *in cis* targeting and fragments C and D, indicating targeting *in trans*. In contrast, the neomycin probe detects only fragment A for an *in cis* targeted clone, and only fragment C for an *in trans* targeted clone.

Fig. 3.2 PFGE blot hybridised with DnLP



The figure shows the results of hybridising restricted DNA from the CAT double-targeted ES cell lines with DnLP. DNA from each line was restricted with either *Not* I alone, or with a combination of *Not* I and then sequentially with *I* *Sce* I. The results were assessed for evidence of a 64 kb fragment arising from the *in cis* targeting of *I* *Sce* I sites within the targeting vectors (see Fig. 3.1). The letters indicating the detected fragments correspond with the annotations used in Fig. 3.1). The fragments that are generated with *Not* I restriction alone are thought to represent fragment B for each of the lines, however for clarity these fragments are not indicated. Only those fragments generated with a combination of *Not* I and *I* *Sce* I have been indicated with their corresponding fragment letter. Clones CAT-A3 and CAT-B3 show the presence of a fragment that corresponds to the fragment between *in cis* targeted *I* *Sce* I sites, i.e. fragment A from Fig. 3.1, indicating that these clones have targeted *in cis*. Clone CAT-C10 demonstrates the absence of fragment A, but the presence of an ~100 kb fragment, thought to represent fragment D from Fig. 3.1, which provided a positive indication of *in trans* targeting for this line. CAT-C6 does not demonstrate the presence of either fragments A or D and therefore presented an ambiguous result with respect to targeting. The fragments that appear to be above 200 kb in size are thought to represent fragment B arising from *in cis* targeted clones, CAT-A3 and CAT-B3. In contrast, an *in trans* targeted clone was predicted to generate fragment C (see Fig. 3.1), as indicated for the CAT-C10 clone.

Fig. 3.3 PFGE blot hybridised with the neomycin probe



The figure shows the results of hybridising restricted DNA from the CAT double-targeted ES cell lines with the neomycin probe. DNA from each line was restricted with either *Not* I alone, or with a combination of *Not* I and then sequentially *I* *Sce* I. The results were assessed for evidence of a 64 kb fragment arising from the *in cis* targeting of *I* *Sce* I sites within the targeting vectors (see Fig. 3.1). The letters indicating the detected fragments correspond with the annotations used in Fig. 3.1. The fragments that are generated with *Not* I restriction alone are thought to represent fragment B for each of the lines, however for clarity these fragments are not indicated. Only those fragments generated with a combination of *Not* I and *I* *Sce* I have been indicated with their corresponding fragment letter. Clones CAT-A3 and CAT-B3 show that the same fragment is detected with the neomycin probe that was detected with DnLP (see Fig. 3.2), which is thought to represent the fragment between *in cis* targeted *I* *Sce* I sites, i.e. fragment A from Fig. 3.1. The detection of this fragment with the neomycin probe thereby provides additional evidence that these clones have targeted *in cis*. Clone CAT-C10 demonstrates the absence of fragment A and, in contrast to Fig. 3.2, the absence of fragment D as well. This demonstrates that the neomycin probe detects only the chromosome targeted with *frt*/I *Sce* I/ 5'hpvt/*loxP*/neomycin and therefore detects only fragment C (as indicated), which is consistent with the assignment of this line as *in trans*. Similarly the fragment originating from the untargeted allele for the *in cis* lines (CAT-A3 and CAT-B3), is not detected with the neomycin probe, as demonstrated by the non-detection of fragment B. This contrasts with Fig. 3.2, since fragment B from the untargeted allele is detected with DnLP. As for the CAT-C6 clone, consistent with Fig. 3.2, the results provide no evidence to establish its targeting arrangement.

The reason that fragment B went undetected following a combination of *Not* I and *I* *Sce* I restriction was due to the inability of the neomycin probe to detect the untargeted chromosome. The neomycin probe detected only the neomycin sequence from the *frit/I* *Sce* I/5'hprr/*loxP*/neomycin upstream targeting vector, which resides on the targeted chromosome (see Fig. 3.1 and Fig. 3.3), therefore since fragment B arises from the untargeted chromosome it is undetected by the neomycin probe. In contrast, the DnLP hybridises with alpha-globin sequences, and is therefore competent to detect both the targeted and untargeted chromosome. Following *Not* I restriction alone, the targeted and untargeted chromosomes generate fragment B, which is detected by DnLP (see Fig. 3.2). However upon using a combination of *Not* I and *I* *Sce* I restriction, the 64 kb *I* *Sce* I fragment (fragment A) is generated from the targeted chromosome and fragment B is generated from the untargeted chromosome and both are detected by the DnLP.

In summary, the correlation of predicted fragment sizes for *in cis* targeted clones to those evident in Figs. 3.2 and 3.3 for clones CAT-A3 and CAT-B3 resulted in their assignment as targeted *in cis*.

The data from Figs. 3.2 and 3.3 were further inspected to determine whether the remaining clones, CAT-B6 and CAT-C10, could be assigned as targeted *in trans*. Fig. 3.1 shows that with *Not* I restriction alone, the DnLP and the neomycin probe were predicted to detect fragment B for the *in trans* targeted clones. As shown in Figs. 3.2 and 3.3, a fragment of this size is present for both the CAT-C6 and CAT-C10 clones with *Not* I restriction alone. However, with respect to the banding patterns evident with a combination of *Not* I and *I* *Sce* I restriction these clones differ.

Fig. 3.1 predicts that the DnLP should detect fragments from *in trans* targeted clones of approximately 210 kb (fragment C) and 90 kb (fragment D) upon a combination of *Not* I and *I* *Sce* I restriction. The data in Fig. 3.2 suggest that these fragments are present for CAT-C10 but not CAT-C6, providing a positive indication of *in trans* targeting for clone CAT-C10, but not CAT-C6. The absence of fragments C and D from the CAT-C6 clone was inexplicable at the time,

however it was anticipated that some error had occurred with the restriction digest or preparation of the DNA from this clone.

The data in Fig. 3.3 further suggested that the CAT-C10 clone had targeted *in trans*. The neomycin probe was competent to detect only the chromosome that had been targeted with the *frt/I Sce* 1/5'hpvt/*loxP*/neomycin targeting vector, as shown in Fig. 3.1. Therefore upon a combination of *Not* I and *I Sce* I restriction the neomycin probe may detect only fragment C. The presence of fragment C and absence of fragment D from clone CAT-C10 was therefore predicted. Fig. 3.3 indeed shows the absence of the 90 kb fragment D and this result further confirmed the assignment of the CAT-C10 line as targeted *in trans*.

The analysis of the CAT-C6 clone provided no evidence to enable its assignment as either *in cis* or *in trans* targeted and so was eliminated from any further investigation in this respect.

The fragment thought to represent fragment A for clones CAT-A3 and CAT-B3 does not appear to be identical for these two clones. The fragment for clone CAT-B3 appears more diffuse. The preparation of DNA for restriction and subsequent PFGE, from ES cells embedded in agarose plugs, includes a stage whereby the plugs are digested with proteases to remove DNA-associated proteins. The incomplete digestion with proteases may affect the motility of the restriction fragments during PFGE due to the persistence of DNA-associated proteins and so for the CAT-B3 cell line this may be the reason that the fragment appears more diffuse.

The PFGE data also showed that fragments B and C appeared to be larger than their predicted size of 240 and 210 kb respectively. The size of fragments B and C were predicted from the position of *Not* I sites identified from the published mouse sequence data (<http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/mapsdb.cgi?org=mouse&chr=11>). However the difference in the position of the *Not* I sites from the published sequence and those experienced for the double-targeted CAT ES cells, that originated from the E14Tg2a line, may represent polymorphisms and thereby explain the difference in fragment sizes experienced. Alternatively, the

parameters for which the PFGE was performed were set to resolve fragments between 10 kb and 100 kb and therefore the size assessment of fragments above this interval may not be accurate.

3.3 Testing whether the hetero-specific *lox* sites in the CAT lines are functional

The analysis described above identified two candidate *in cis* targeted lines to participate in the large interval RMCE experiment, namely CAT-A3 and CAT-B3. CAT-C10 was assigned as targeted *in trans*. These lines were further tested to establish whether the *lox* sites they harboured were functional.

In order to test the functionality of the *lox* sites, their ability to undergo Cre-mediated SSR with the test plasmid, CP#98, which is shown in Fig. 3.4, was investigated. CP#98 includes the 3' region of the *hprt* gene, which upon SSR with the *loxP* site in the *frt/I Sce I/5'hprt/loxP/neomycin* upstream targeted construct is capable of reconstructing a functional *hprt* gene. Additionally CP#98 includes a *lox511* site, which was included to investigate whether the efficiency of reconstructing the *hprt* gene in the *in cis* targeted clones would be different to that of the *in trans* targeted clone.

The *loxP* and *lox511* sites in CP#98 were orientated such that upon SSR with an *in cis* targeted clone it would be possible for the 3' region of the *hprt* gene to replace the interval that encompasses the alpha-globin locus, as illustrated in Fig. 3.5a. In contrast, the replacement of the alpha-globin locus in an *in trans* targeted clone by SSR with CP#98 should not be possible. However, it was anticipated that HAT^R clones could be generated from an *in trans* targeted clone by Cre-mediated insertion of the entire CP#98 plasmid via its *loxP* site into the *loxP* site of the *frt/I Sce I/5'hprt/loxP/neomycin* upstream targeted construct. This event would reconstruct the *hprt* minigene and provide HAT^R, as illustrated in Fig. 3.6a. It is important to note however, that the *in cis* targeted clones could also undergo insertion of CP#98 to generate HAT^R clones, and more significantly this would not be concomitant with replacement of the alpha-globin interval.

Fig. 3.4 Construction of CP#98, a plasmid used to test the functionality of the *lox* sites in the double-targeted CAT ES cell lines

Oligos W2186 and W2187 were designed to include an *frt* site, an I *Sce* I site and *Bam*HI compatible ends. They were annealed and then ligated into AJHS#924 (see Fig. 3.4a), which had been previously restricted with *Bam*HI. The resulting colonies were screened by colony lift, using ³²P labelled W2187 as a probe. Positive clones were sequenced with T7 and T3 primers to establish the orientation of the oligos. The plasmid containing the desired orientation of oligos W2186/W2187 was termed CP#51 (see Fig. 3.4 b) and harboured the following arrangement: *loxP* – 3'HPRT – *Bam*HI destroyed – I *Sce* I – *frt* – *Bam*HI. Oligos F22433 and F22434 were designed to introduce a *lox511* site between the *Bam*HI and *Kpn*I site in CP#51. By designing the oligos with one end compatible to *Bam*HI and the other to *Kpn*I the orientation of the inserted oligos was assured. Additionally, the oligos were designed to introduce a unique *Nco*I site, which was subsequently used to screen for their presence. CP#51 was *Bam*HI and then *Kpn*I restricted and the annealed F22433/F22434 oligos inserted. The oligos destroyed the original *Bam*HI and *Kpn*I sites and introduced a new *Bam*HI site to create a cassette that could be isolated by *Bam*HI restriction alone. The *Nco*I site was used to screen the resulting colonies, however positive plasmids were additionally sequenced to verify the integrity of the *lox511* site. The desired arrangement was: *loxP* – 3'HPRT – *Bam*HI destroyed – I *Sce* I – *frt* – *Bam*HI destroyed – *lox511* – *Nco*I – *Bam*HI – *Kpn*I destroyed (see Fig. 3.4c). The *lox* sites were arranged in an inverted orientation, consistent with the hetero-specific *lox* sites targeted to the CAT ES cell lines, which would therefore permit the deletion of the entire alpha-globin locus from an *in cis* targeted line upon Cre-mediated SSR.

Fig. 3.4 Construction of CP#98, a plasmid used to test the functionality of the *lox* sites in the double-targeted CAT ES cell lines

Fig. 3.4a

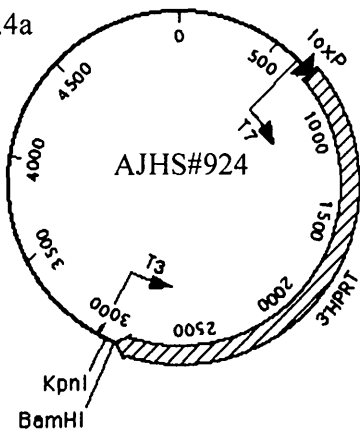


Fig. 3.4b

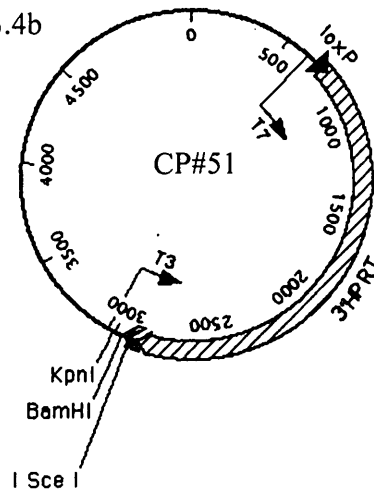


Fig. 3.4c

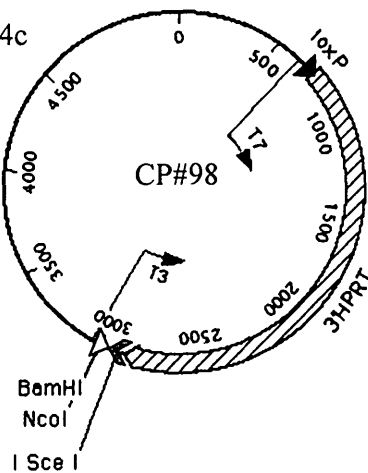


Fig. 3.5 Functional testing of the *lox* sites in the recombination cassettes by SSR between CP#98 and an *in cis* targeted clone

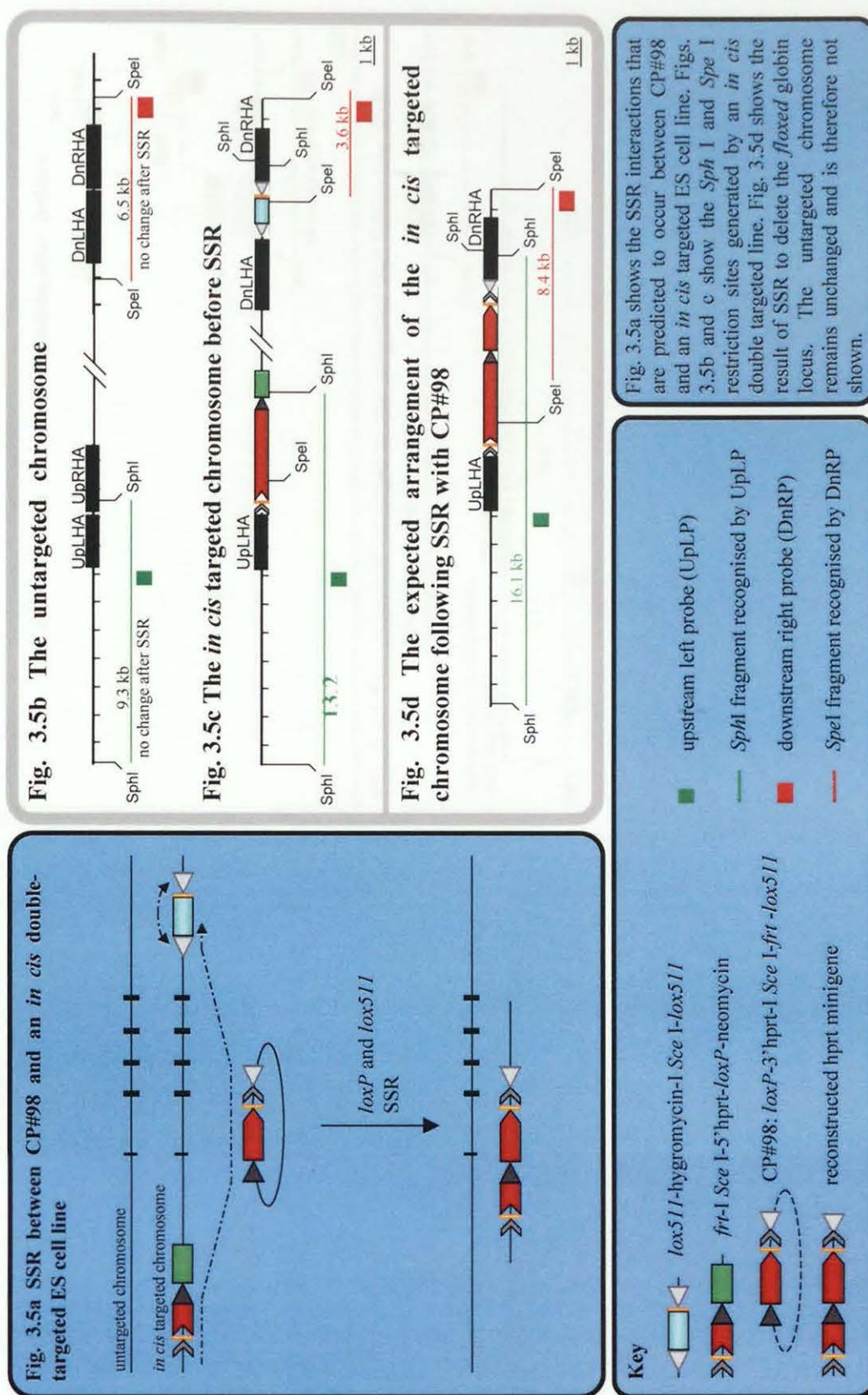


Fig. 3.6 Functional testing of the *lox* sites in the recombination cassettes by SSR between CP#98 and an *in trans* targeted clone

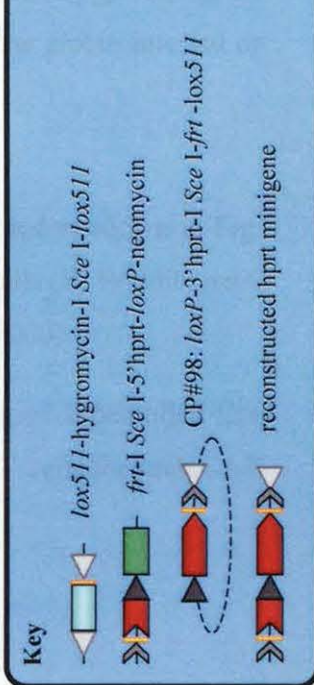
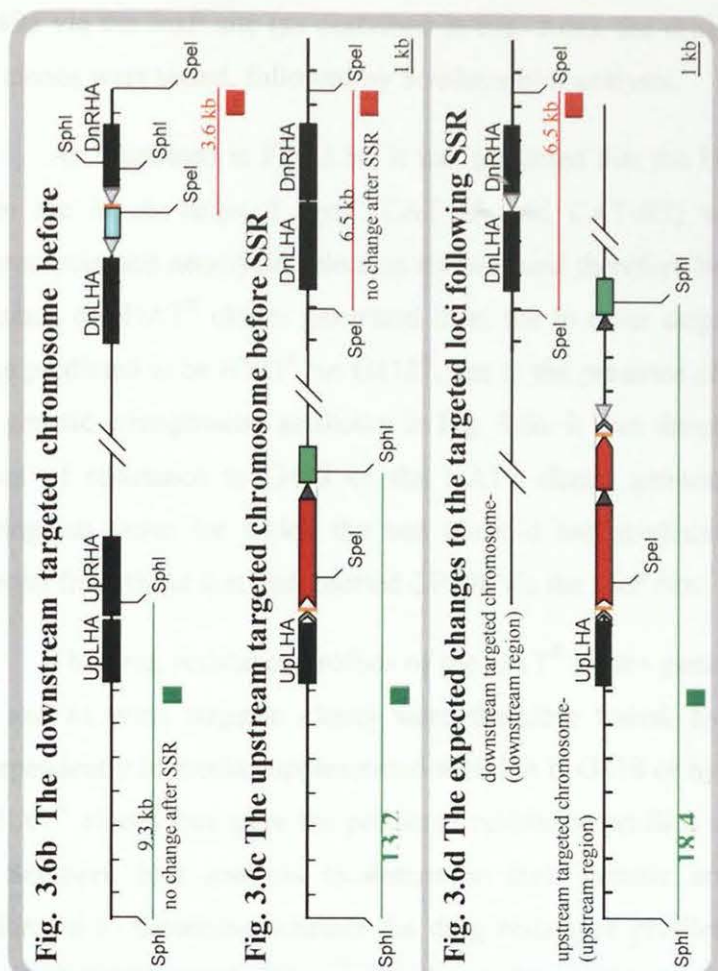
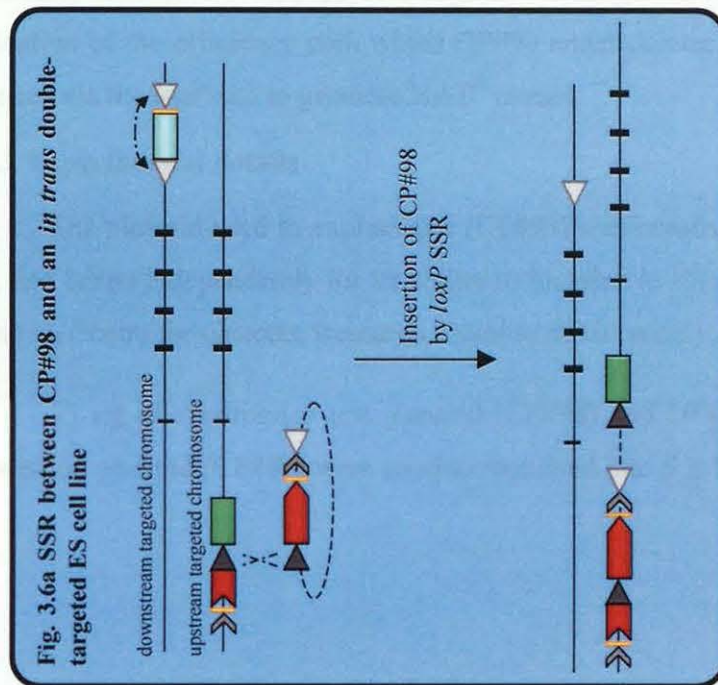


Fig. 3.6a shows the SSR interactions that are predicted to occur between CP#98 and an *in trans* targeted ES cell line. Figs. 3.6b and c show the *SphI* and *SpeI* restriction sites generated by an *in trans* double targeted line. Fig. 3.6d shows the changes in restriction fragment sizes following SSR, the untargeted loci remain unchanged and are therefore omitted.

■ upstream left probe (UpLP)

— *SphI* fragment recognised by UpLP

■ downstream right probe (DnRP)

— *SpeI* fragment recognised by DnRP

To distinguish HAT^R clones arising from the deletion of the hetero-specifically *floxed* interval (as described in Fig. 3.5a) from those that had inserted CP#98 via the *loxP* site (as described in Fig. 3.6a), the drug resistance profiles of the clones were tested, followed by Southern blot analysis.

As illustrated in Fig. 3.5a, it was predicted that the HAT^R clones generated from the *in cis* targeted lines (CAT-A3 and CAT-B3) would delete both the hygromycin and neomycin selection markers and therefore be HYG^S and G418^S. In contrast, the HAT^R clones generated from the *in trans* targeted clone (CAT-C10) were predicted to be HYG^S but G418^R, due to the presence of the neomycin gene in the genetic arrangement, as shown in Fig. 3.6a. It was therefore predicted that the status of resistance to G418 of the HAT^R clones generated could be used to distinguish those for which the test plasmid had mediated the deletion of the interval from those that had inserted CP#98 via the *loxP* site.

The drug resistance profiles of the HAT^R clones generated from both the *in cis* and *in trans* targeted clones were therefore tested, by selecting the clones independently in media supplemented with HAT, G418 or hygromycin. A selection of HAT^R clones that gave the predicted resistance profiles were then investigated by Southern blot analysis to determine their genetic arrangement. This was performed to determine whether the drug resistance profiles of the HAT^R clones accurately represented their predicted genetic arrangement, thereby providing an indication of the efficiency with which CP#98 either deleted the globin interval or inserted via the *loxP* site to generate HAT^R clones.

3.3.1 Experimental details

The plasmid used to express Cre (CP#99) was constructed as shown in Fig. 3.7. and tested independently for its ability to function in ES cells (H. Priddle, pers. comms., Centre for Genome Research, Edinburgh University, 2000).

75 µg of supercoiled test plasmid (CP#98) and 50 µg of supercoiled Cre expression plasmid (CP#99) were co-electroporated into 5×10^7 cells for each CAT

Fig. 3.7 Construction of the Cre expression plasmid, CP#99

A plasmid that could express Cre constitutively within ES cells was generated in order to test the function of the *lox* sites within the CAT lines and to provide Cre for the desirable large interval RMCE experiment. The CreEBD gene fusion sequences were PCR amplified with *Pfu* Turbo from CP#93 (pNPKCreERc) with primers B8141 and B8146 (see Fig. 3.7a). B8141 and B8146 included sequences to generate *Eco*RI sites at either end of the resulting PCR fragment. The PCR product, containing the CreEBD sequences, was therefore *Eco*RI restricted and then ligated into *Eco*RI restricted CP#15 (AS#575 pCAGSIP). The desirable orientation of the insert was determined by *Bam*HI restriction and the resulting plasmid termed CP#64 (see Fig. 3.7b). The integrity of the CreEBD sequence, following PCR amplification with *Pfu* Turbo, was determined by sequencing the entire ~ 2kb fragment with a series of overlapping primers (see Appendix III) and generating a contig alignment of the resulting sequence data. This sequence was then compared to the published sequence from F. Stewart's web page (see Appendix III) to ensure that no mutations had been introduced during the PCR that may compromise the activity of the CreEBD gene fusion (data not shown). To produce a constitutively expressed Cre recombinase gene the EBD region of the gene fusion was removed (see Fig. 3.7c) by restriction. This was achieved by partial restriction with *Eco*RI, followed by complete restriction with *Bsi*WI and then end-filling and re-ligation (performed by Helen Priddle).

Fig. 3.7 Construction of the Cre expression plasmid, CP#99

Fig. 3.7a

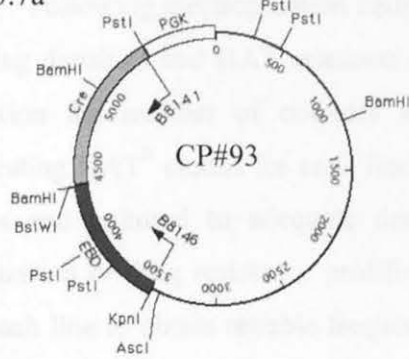


Fig. 3.7b

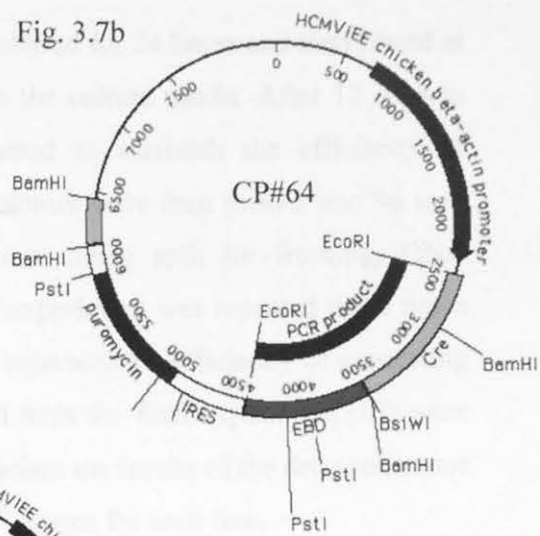
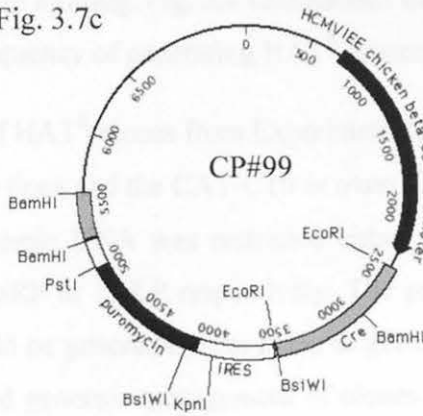


Fig. 3.7c



cell line tested. The *in cis* lines chosen for test were CAT-A3 and CAT-B3 and the *in trans* line was CAT-C10.

Following electroporation cells were cultured for 24 hours and then plated at varying densities and HAT selection added to the culture media. After 12 days in selection the number of colonies was counted to establish the efficiency of generating HAT^R clones for each line. The colonies were then picked into 96 well plates and cultured to adequate density before being split for freezing, DNA preparation or drug resistance profiling. This experiment was repeated three times for each line to obtain reliable frequencies to represent the efficiency of generating HAT^R clones. A selection of clones generated from the final experiment (III) were analysed by Southern blotting. Fig. 3.8 summarises the results of the drug resistance profiles and the frequency of generating HAT^R clones for each line.

A sample of HAT^R clones from Experiment III, generated from the CAT-A3 and CAT-B3 *in cis* lines and the CAT-C10 *in trans* line, were selected for Southern blot analysis. Genomic DNA was restricted either with *Spe* I or *Sph* I and then hybridised with DnRP or UpLP respectively. The predicted genomic arrangement of clones that would be generated from *in cis* targeted lines is illustrated in Fig. 3.5, whilst the predicted genomic arrangement of clones that would be generated from *in trans* targeted clones is illustrated in Fig. 3.6. The results of the Southern blot analysis are shown in Fig. 3.9 and Fig. 3.10, for clones generated from either the *in cis* or *in trans* lines respectively.

3.4 Interpretation of the results from testing the *lox* site functionality

The results generated from co-electroporating selected CAT lines with CP#98 and CP#99 were assessed to determine whether they verified the assignment of these clones as targeted *in cis* or *in trans*, as determined by the analysis based upon I *Sce* I restriction. Additionally it was assessed whether this strategy could be used to distinguish *in cis* from *in trans* clones based upon their resistance to G418 supplemented media.

Fig. 3.8 Summary of the results from electroporation of selected CAT lines with CP#98 and CP#99

Experiment number	CAT line	Efficiency of generating HAT ^R clones	Number of clones with the expected drug resistance phenotype (*)	Number of clones providing no results	Number of clones with unexpected drug resistance phenotypes	Details of unexpected drug resistance phenotypes
I	CAT-A3	2.1 x 10 ⁻⁴	23 / 24	0 / 24	1 / 24	Clone A7: HAT ^R /G418 ^S /HYG ^R
	CAT-B3	2.1 x 10 ⁻⁴	26 / 26	0 / 26	0 / 26	
	CAT-C10	0.4 x 10 ⁻⁴	0 / 4	0 / 4	4 / 4	Clones D1, D2, D3 and D4: HAT ^R /G418 ^S /HYG ^S
II	CAT-A3	4.8 x 10 ⁻⁴	Not tested due to contamination			
	CAT-B3	2.5 x 10 ⁻⁴				
	CAT-C10	0.6 x 10 ⁻⁴				
III	CAT-A3	1.5 x 10 ⁻⁴	8 / 13	5 / 13	0	
	CAT-B3	5.0 x 10 ⁻⁴	80 / 96	18 / 96	2 / 96	Clone C6: HAT ^R /G418 ^R /HYG ^S Clone F12: HAT ^R /G418 ^S /HYG ^R
	CAT-C10	0.8 x 10 ⁻⁴	0 / 1	0 / 1	1 / 1	Clone A1: HAT ^R /G418 ^S /HYG ^S

The table summarises the results from three separately conducted experiments, each using the same experimental parameters. (*) The expected drug resistance phenotype for the CAT lines putatively assigned as *in cis* targeted (CAT-A3 and CAT-B3) was HAT^R/G418^S/HYG^S, whilst for the *in trans* targeted line (CAT-C10), it was expected to be HAT^R/G418^R/HYG^S. The last column details the profiles of clones with unexpected drug resistance phenotypes, with the unexpected phenotype result highlighted in bold.

3.4.1 Comparing the efficiency of generating HAT^R clones between *in cis* targeted and *in trans* targeted lines

The data in Fig. 3.8 show that for each independent experiment the lines assigned as *in cis* targeted (CAT-A3 and CAT-B3) consistently generated HAT^R colonies at a higher frequency than for the CAT line that had been assigned as targeted *in trans* (CAT-C10). The reliability of these frequencies was based upon the consistency with which HAT^R clones were generated, which were at frequencies of between 2 to 5 colonies per 1×10^4 cells (with the exception of the CAT-A3 line for Experiment III, see below). In addition, the frequency of generating a HAT^R colony from the CAT-C10 clone was consistently less than 1 per 1×10^4 cells.

The cause of the reduced efficiency with which the CAT-A3 line generated HAT^R clones for Experiment III was not definitively established, however it was noted that these cells were multiplying at a slower rate than for the CAT-B3 line, which may be attributed to a mycoplasma contamination of the cells during this experiment. A contamination of the CAT-A3 line in this manner may have reduced their efficiency to survive HAT selection, thereby reducing the efficiency with which HAT^R colonies were generated.

The difference in the efficiencies of generating HAT^R clones between the *in cis* and *in trans* lines had been anticipated based upon the reported findings of Bethke and Sauer (Bethke and Sauer, 1997). Bethke and Sauer conducted an experiment to determine whether the insertion of an entire plasmid by a single SSR event (see Fig. 1.2c) would occur with the same efficiency as an RMCE event (see Fig. 1.2e). In the experiment, the target plasmid was the same harbouring both *loxP* and *lox511* sites, between which was an interval of approximately 3 kb. However, the test plasmids differed: one contained the equivalent *loxP* and *lox511* sites, creating an interval of approximately 4 kb, which was used to test the efficiency of the RMCE event; the other test plasmid was identical in all respects, except that it contained just a *loxP* site and was therefore used to test the efficiency with which the plasmid inserted into the target. Bethke and Sauer found that the RMCE event,

using the plasmid with *loxP* and *lox511*, was recovered with a 20-fold greater efficiency than the insertion event, using the plasmid with *loxP* alone. They argued that the reason for the difference in the efficiencies was based upon the likelihood for the reverse SSR event occurring in each case. They reasoned that the insertion of the test plasmid, with the *loxP* site alone, by a single SSR event would occur as efficiently as its excision from the target, again by a single SSR event. However, the implementation of hetero-specific *lox* sites, with the RMCE test plasmid, provided the opportunity for a different outcome: following insertion of the RMCE test plasmid via either *loxP* or *lox511*, a second SSR event could mediate the reverse excisive reaction, however the possibility also existed that the second SSR could occur between the other pair of *lox* sites and thereby generate a RMCE event. Furthermore, Bethke and Sauer argued that the reverse reaction to RMCE would be an extremely unlikely event. This was based upon their supposition that the RMCE event was reciprocal and would therefore generate a plasmid that harboured the target interval of 3 kb. In order to reverse the RMCE event, this reciprocal molecule would then have to compete with 'a vast excess' of the original targeting plasmid and Bethke and Sauer considered the likelihood that this would occur to be extremely low. Therefore the reverse reaction for RMCE was reasoned to be extremely unlikely, whilst integration and its reverse reaction of excision, were considered to be of equal likelihood.

The consequences of altering the exposure time to Cre activity on integration/excision systems should therefore be less influential on determining their SSR products, compared with a RMCE system. The integration/excision system would be predicted to reach equilibrium of inserted and excised genetic products, in equal proportions, which would then become stabilised in the absence of Cre. However, since the reverse reaction to RMCE is predicted to be extremely unlikely, the prolonged exposure of Cre within a RMCE system would therefore be predicted to favour the generation RMCE events. Therefore RMCE systems present a greater capacity to increase the yield of desired products compared with integration/excision systems.

With respect to the experiments described in this section, the production of HAT^R clones from the *in trans* line was predicted to follow equivalent dynamics to the integration of the plasmid by a single SSR event via the *loxP* site. The deletion of the interval was predicted to be equivalent to the RMCE event, outlined by Bethke and Sauer, due to the greater likelihood of a double SSR event, involving both the *loxP* and *lox511* sites. Therefore, in accordance with the findings of Bethke and Sauer the deletion of the globin locus, in the *in cis* lines, appears to be more efficient than the generation of HAT^R clones by integration of CP#98 into the *in trans* line.

It was predicted therefore that the HAT^R clones that were generated from the CAT-C10 line would arise from the integration of the entire CP#98 plasmid via its *loxP* site into the *loxP* site within the *frit*/ I *Sce* I/5'hppt/*loxP*/neo upstream targeting construct and that their drug resistance profiles would therefore reflect this arrangement. Similarly the HAT^R clones from CAT-A3 and CAT-B3 were predicted to have drug resistance profiles that reflected the deletion of the interval encompassing the entire alpha-globin locus.

3.4.2 Drug resistance phenotypes from HAT^R clones generated from the *in cis* and *in trans* targeted lines

As shown in Fig. 3.8 the drug resistance phenotypes for CAT-A3, CAT-B3 and CAT-C10 were determined for Experiments I and III, but not for Experiment II, since contamination prevented this assessment.

The data in Fig. 3.8 show that the majority of the HAT^R clones generated from the CAT-A3 and CAT-B3 lines had drug resistance phenotypes that were consistent with the predicted phenotypes of clones that would be generated from *in cis* targeted lines (i.e. HAT^R, G418^S and HYG^S). In fact, of those clones that gave a drug resistance phenotype, over 90 % had drug resistance phenotypes consistent with being generated from *in cis* targeted lines. This was taken as a strong indication that these clones were targeted *in cis*.

Those clones that gave unexpected drug resistance profiles from the putative *in cis* lines were clones A7 (Experiment I) and F12 (Experiment III), which were

unusually hygromycin resistant and clone C6 (Experiment III), which was unusually G418 resistant. Clones A7 and F12 could be explained by the absence of Cre recombinase from these clones, which would be necessary to delete the *floxed* hygromycin gene. However, the sensitivity of both these clones to G418 and their HAT^R status, would suggest that the Cre-mediated deletion of the alpha-globin locus had occurred within these clones, as illustrated in Fig. 3.4. It was therefore probable that Cre had been present and had mediated the deletion of 64 kb interval of the globin locus, but surprisingly not an interval of 1.8 kb to delete the hygromycin selection marker. However, the occurrence of these events at such a low frequency (5×10^{-6}) negated their relevance to the assignment of *in cis* targeting and therefore excluded them from further investigation.

Clone C6 is unusually G418^R, but HAT^R and HYG^S, which would suggest that the entire CP#98 plasmid has inserted via the *loxP* site into the *loxP* site adjacent to 5'hpvt, in an SSR event similar to that predicted to generate HAT^R clones from the *in trans* line, see Fig. 3.6. According to the data in Fig. 3.8, this event occurs at a very low frequency, at 5×10^{-6} . The verification of the genetic arrangement of clone C6 requires Southern blot analysis, however, if this clone does indeed represent an integration of CP#98, via *loxP*, then it appears that the deletion of the globin locus interval via the SSR event outlined in Fig. 3.5, occurs 10^2 times more efficiently than the integration event, for *in cis* targeted lines. This is further consistent with the findings of Bethke and Sauer, as outlined above.

The generation of clone C6 (Experiment III) at a frequency of 5×10^{-6} would suggest that the generation of HAT^R clones from the *in trans* line should occur with equivalent frequency, since they are generated from equivalent SSR events. However, although the frequency of generating HAT^R clones from the *in trans* line was lower than for the *in cis* lines, the frequency was significantly higher than 5×10^{-6} , being 6×10^{-5} . This was not the only anomaly associated with the HAT^R clones generated from the *in trans* line. All of the HAT^R clones were sensitive to G418 supplemented media. The predicted genetic arrangement of HAT^R clones from the *in trans* line is shown in Fig. 3.6, which illustrates the persistence of the neomycin gene, thereby conferring resistance to G418

supplemented media. The sensitivity of the HAT^R clones generated from the *in trans* lines to G418 was therefore confusing.

The drug resistance profiles for the HAT^R clones generated from the CAT-A3 and CAT-B3 lines were consistent with their assignment as targeted *in cis*, however Southern blot analysis was performed on a selected number of clones to further verify this conclusion (see below). Clearly, the drug resistance profiles of the HAT^R clones generated from the CAT-C10 clone were not consistent with the predicted genetic arrangement following Cre-mediated SSR with CP#98, which is illustrated in Fig. 3.6, and it was hoped that Southern blot analysis would clarify the assessment of this line.

3.4.3 Interpretation of the Southern blot data from the HAT^R clone generated from the CAT lines assigned as targeted *in cis*

Eight HAT^R clones generated from the CAT-A3 clone and ten HAT^R clones from the CAT-B3 clone, that had the predicted drug resistance phenotype of HAT^R, G418^S and HYG^S, underwent Southern blot analysis to determine their genetic arrangement (see Fig. 3.9). Upon *Sph* I restriction and hybridisation with UpLP, the control untargeted ES cell DNA demonstrates the presence of a 9.3 kb fragment, whilst the CAT-B3 double-targeted line shows this fragment and an additional fragment of 13.2 kb (as predicted in Fig. 3.5b and c). Following SSR the 13.2 kb fragment should be absent, since the *Sph* I site within the neomycin gene should be removed upon deletion of this selection marker concomitant with the predicted deletion of the hetero-specifically *floxed* interval. Furthermore a fragment of 16.1 kb should be generated by the occurrence of an *Sph* I site within the DnRHA (see Fig. 3.5d). This fragment is particularly evident of a deletion, more so than a fragment which would be created by a site in the 3'hpvt region. This is because it is contingent on the *Sph* I site in the DnRHA being brought into proximity with the reconstructed hpvt gene, which would only occur by the deletion of the entire interval as predicted. All of the clones generated from both CAT-A3 and CAT-B3 demonstrate the presence of this 16.1 kb fragment, in addition to the 9.3 kb

Fig. 3.9 Southern blot analysis from a selected number of HAT^R clones generated from the co-electroporation of CP#98 and CP#99 into CAT-A3 and CAT-B3 ES cell lines (Experiment III)

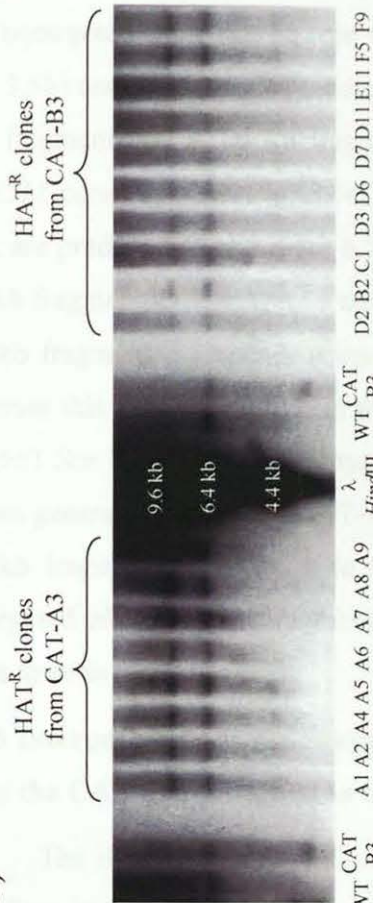


Fig. 3.9a DNA from selected HAT^R clones restricted with *Spe* I and hybridised with DnRP

Fig. 3.9b DNA from selected HAT^R clones restricted with *Sph* I and hybridised with UpLP

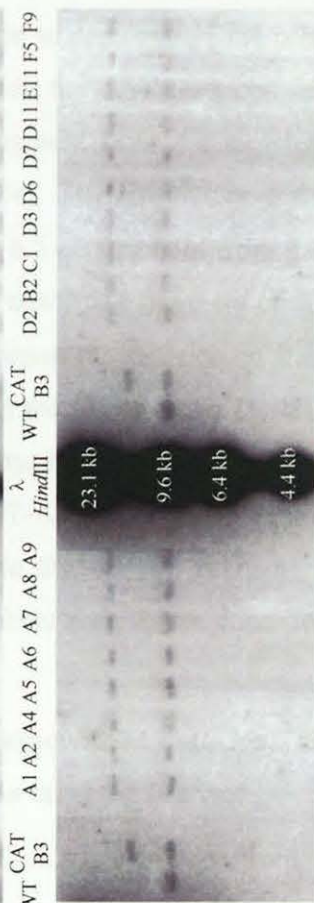


Fig. 3.9 shows the fragments detected when DNA from a selected number of HAT^R clones arising from the electroporation of putatively assigned *in cis* lines were co-electroporated with CP#98 and CP#99. WT indicates DNA from un-targeted E14Tg2a cells and as predicted from Fig. 3.5 generates fragments of 6.5 kb following *Spe* I restriction, detected by DnRP, as shown in Fig. 3.9a. This fragment is reduced in size to 3.6 kb following targeting of the *lox511/hygro* construct (as shown for the CAT-B3 line) and increased to 8.4 kb following deletion of the *alpha*-globin interval by SSR. As shown in Fig. 3.9a, all of the CAT-A3 and CAT-B3 generated clones produce the predicted fragment sizes, verifying these lines as targeted *in cis*. Fig. 3.9b shows the fragments detected when DNA is restricted with *Sph* I and hybridised with UpLP, which are illustrated in Fig. 3.5. The untargeted locus should generate a fragment of 9.3 kb (see WT lane), which is increased in size to 13.2 kb following targeting of the *frt/1 Sce 1/5'hpvt/loxP/neomycin* construct (as shown for the CAT-B3 line) and then further increased to 16.2 kb following deletion of the *alpha*-globin interval by SSR. As shown, all of the CAT-A3 and CAT-B3 clones generate the predicted fragment sizes, which further verified the CAT-A3 and CAT-B3 lines as targeted *in cis*.

fragment, which originates from the untargeted chromosome. These fragments are therefore consistent with the assignment of these lines as targeted *in cis*.

The conclusions drawn following the interpretation of the data for genomic DNA restricted with *Spe* I and hybridised with DnRP, are similar to those above. The untargeted DNA can be seen to demonstrate the predicted 6.5 kb fragment (see Fig. 3.5b) and following targeting the CAT-B3 clone demonstrates the presence of this fragment and a 3.6 kb fragment from the targeted *lox511/hygromycin/I Sce I/lox511* construct (see Fig. 3.5c). The HAT^R clones that are generated following SSR are predicted to retain the 6.5 kb fragment from the untargeted allele, lose the 3.6 kb fragment and generate a new *Spe* I fragment of 8.4 kb. The generation of the 8.4 kb fragment is dependent upon the deletion of the interval since the sites to generate this fragment reside in the region 3' of DnRHA and within the 5'hprt of the *frt/I Sce I/5'hprt/loxP/neomycin* targeting construct (see Fig. 3.5d). All of the clones generated from both CAT-A3 and CAT-B3 demonstrate the presence of this 8.4 kb fragment, in addition to the 6.5 kb fragment which originates from the untargeted allele. These fragments are further consistent with the assignment of these lines as targeted *in cis*.

3.4.3 Interpretation of the Southern blot data from the HAT^R clone generated from the CAT lines assigned as targeted *in trans*

The HAT^R clone CAT-C10/A1, generated from the CAT-C10 (*in trans*) ES cell line in Experiment III, was assessed by Southern blotting. The predicted fragment changes are illustrated in Figs. 3.6b, c and d, whilst the results are shown in Fig. 3.10.

Fig. 3.6 predicts that the hygromycin gene will be deleted following the exposure of the CAT-C10 *in trans* line to Cre recombinase. This causes a concomitant change in the size of the *Spe* I fragment from the targeted chromosome, from 3.6 kb to 6.5 kb. In addition, the untargeted chromosome was predicted to remain unchanged and should also generate a *Spe* I fragment of 6.5 kb. The CAT-C10/A1 clone clearly shows the presence of the 6.5 kb fragment and the

Fig. 3.10 Southern blot analysis from the HAT^R clone generated from the co-electroporation of CP#98 and CP#99 into the CAT-C10 ES cell line (Experiment III)

Fig. 3.10a DNA from the HAT^R clone restricted with *Spe* I and hybridised with DnRP

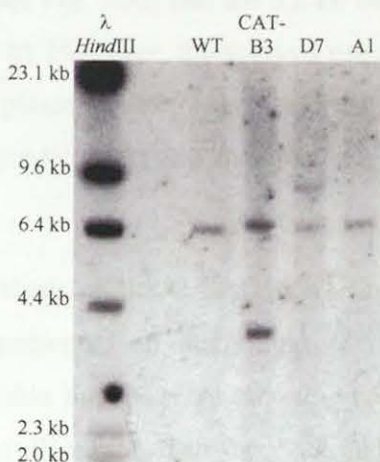


Fig. 3.10b DNA from the HAT^R clone restricted with *Sph* I and hybridised with UpLP

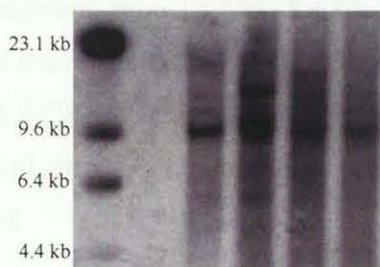


Fig. 3.10 shows the fragments detected when DNA from the HAT^R clone arising from the electroporation of the putatively assigned *in trans* line was co-electroporated with CP#98 and CP#99. WT indicates DNA from un-targeted E14Tg2a cells, CAT-B3 is the control double-targeted line, clone D7 is a HAT^R clone generated from the CAT-B3 line and was assessed as having undergone deletion of the alpha-globin interval (see Fig. 3.9, clone D7) and clone A1 is generated from the CAT-C10 clone. The CAT-C10/A1 clone is thought to represent a HAT^R clone generated by the insertion of the test plasmid CP#98, not concomitant with deletion of the alpha-globin interval (see Fig. 3.6 for the predicted fragment sizes). As predicted from Fig. 3.6b, WT DNA generates fragments of 6.5 kb following *Spe*I restriction, detected by DnRP, as shown in Fig. 3.10a. This fragment is reduced in size to 3.6 kb following targeting of the *lox511/hygromycin/1Sce I/lox511* construct (as shown for the CAT-B3 line) and increased back to 6.5 kb following deletion of the hygromycin gene (see Fig. 6.5d). As shown in Fig. 3.10a, the CAT-C10/A1 clone shows the fragment sizes consistent with the deletion of the hygromycin gene and additionally does not demonstrate the presence of an 8.4 kb that is indicative of *in cis* derived HAT^R clones (see clone D7 above, as predicted in Fig. 3.5d). Fig. 3.10b shows the fragments detected when DNA is restricted with *Sph*I and hybridised with UpLP, which are predicted in Fig. 3.6. The untargeted locus should generate a fragment of 9.3 kb (see WT lane), which is increased in size to 13.2 kb following targeting of the *frt1Sce I/5'hprt/loxP/neomycin* construct (as shown for the CAT-B3 line). The D7 control line was predicted to generate a 16.2 kb (see Fig. 3.5d), indicative of deletion of the alpha-globin interval by SSR, however this fragment is not evident, as it was in Fig. 3.9b. Additionally, a fragment of 18.4 kb that would indicate the insertion of the test plasmid CP#98 (see Fig. 3.6d) is not evident for the CAT-C10/A1 clone. Therefore the results for the *Sph*I restriction on the HAT^R clone generated from the *in trans* line do not confirm the predicted genomic arrangement described in Fig. 3.6d.

absence of the 3.6 kb fragment, which is consistent with the deletion of the hygromycin gene, and furthermore demonstrates that Cre recombinase has been active in this clone. Additionally, the CAT-B3/D7 clone, which is included as a control, demonstrates the presence of the 8.4 kb *Spe* I fragment which arises following deletion of the alpha-globin locus (see Fig. 3.5d) and the 6.5 kb fragment from the untargeted chromosome (see Fig. 3.5 b). However, the analysis using *Spe* I restriction does not indicate whether the test plasmid, CP#98, has inserted via the *loxP* site to reconstruct the Hprt minigene, as predicted for this *in trans* line. To test this *Sph* I restriction was performed.

The results of the *Sph* I restriction analysis on the CAT-C10/A1 line were predicted to indicate whether this clone represented an integration of the test plasmid via its *loxP* site, into the *loxP* site within the upstream targeting construct (as shown in Fig. 3.6a). Figs. 3.6c and d illustrate that there should be a concomitant increase in the size of the *Sph* I fragment, from 13.2 to 18.4 kb, around the hprt minigene should integration occur. As the results in Fig. 3.10 demonstrate, this 18.4 kb fragment is not evident in the CAT-C10/A1 clone. The only fragment that is evident is that arising from the untargeted chromosome, of 9.3 kb (as shown in Fig. 3.6b). Clearly this is an anomalous result.

The results shown in Fig. 3.10, of *Sph* I restriction on the E14Tg2a (WT) and CAT-B3 control lines demonstrated the predicted fragments. However the results from the CAT-B3/D7 control clone are not in accordance with its predicted genetic arrangement, or the previous assessment of this line (see Fig. 3.9b). The data previously generated from this clone indicated that it had undergone deletion of the alpha-globin locus (see Fig. 3.5b), as predicted (see Fig. 3.5). However, in the analysis presented alongside the CAT-C10/A1 clone, in Fig. 3.10, the CAT-B3/D7 clone does not demonstrate the presence of the 16.1 kb fragment, which is indicative of the deletion event (Fig. 3.5), however the 9.3 kb fragment from the untargeted chromosome is detected. Taken together, the results for the CAT-C10/A1 clone and CAT-B3/D7 clone, in Fig. 3.10, would suggest that the *Sph* I

analysis performed does not provide conclusive data that either verify, or exclude, the assignment of the CAT-C10 clone as targeted *in trans*.

3.5 Summary

A selection of four double-targeted CAT ES cell lines were further characterised to determine whether the targeting constructs, whose targeting was described in Chapter 2, had targeted to the same chromosome (*in cis*) or alternate chromosomes (*in trans*). The initial analysis was to perform Southern blot analysis on genomic DNA from these lines restricted with I *Sce* I, which is a rare-cutting restriction endonuclease and whose sites were exclusively present within the targeting constructs. Therefore the generation of an I *Sce* I restriction fragment was exclusively dependent upon the constructs targeting *in cis*. The detection of the I *Sce* I fragment by Southern blot analysis, following PFGE, was used to putatively assign the CAT-A3 and CAT-B3 lines as targeted *in cis* and the CAT-C10 line as targeted *in trans*.

Following the putative assignment of the CAT lines in this manner, they were further characterised to verify that the *lox* sites they harboured, within the targeted constructs, were functional. This was achieved by co-electroporating a Cre expression plasmid and a test plasmid into the three CAT lines. The test plasmid contained *loxP* and *lox511* sites, in an orientation that would permit the deletion of the entire alpha-globin locus, upon SSR with an *in cis* targeted clone. This analysis further verified the assignment of clones CAT-A3 and CAT-B3 as targeted *in cis*. This analysis did not however conclusively verify the assignment of the CAT-C10 clone as targeted *in trans*.

The CAT-B3 line was selected to attempt the large interval RMCE experiment, which is described in Chapter 5. The CAT-A3 line was not selected due to its slow duplication rate experienced for this clone in Experiment III (see section 3.4.1).

Although the CAT-C10 clone generated data that contradicted its assignment as targeted *in trans*, it was still used as a comparative cell line for the

large interval RMCE event, described in Chapter 5. It was hoped that the CAT-C10 line would generate interesting data, for comparison with the CAT-B3 line, which was selected to achieve the large interval RMCE event proper.

Chapter 4

BAC TARGETING

4.1 Introduction

This chapter describes the genetic manipulation of the bacterial artificial chromosome pBAC14567 in order to modify it in preparation for the large interval RMCE event, as described briefly in section 1.5.4(a) (see Fig. 1.3). pBAC14567 consists of the BAC cloning vector pBeloBAC11 and a *HindIII* cloned insert of approximately 70 kb of murine 129 genomic DNA. The cloned insert in pBAC14567 encompasses the murine alpha-globin locus, commencing 1 kb 5' of the HS-26 sequence and terminating 10.2 kb 3' of last coding exon of θ , which is the terminal structural gene of the alpha-globin locus gene cluster (see Fig. 2.1).

The modification of pBAC14567 involved targeting the cloned insert with hetero-specific *lox* sites, at positions that were equivalent to those targeted during the modification of the CAT ES cell lines, as described in Chapter 2. The cloned insert in pBAC14567 was targeted by two distinct sequential targeting events, using recombination mediated cloning. The first targeting event delivered a construct, containing *lox511* sites, targeting 3' of the alpha-globin locus at the 'downstream' position in the pBAC14567 (see Fig. 2.1 for nomenclature). In addition this recombination cassette contained a Herpes Simplex Virus thymidine kinase gene (HSV ι k). The isolation of clones targeted with this recombination cassette was aided by the inclusion of a bacterial kanamycin (kan) drug resistance gene as a

selectable marker. The design and construction of this recombination cassette, termed *lox511/kan/lox511/HSVtk*, is described in section 4.2.

Targeting of the recombination cassettes to pBAC14567 was achieved by recombination mediated targeting (Zhang et al., 2000). Before a successful method was established a number of different approaches were attempted in order to generate the double-stranded linear targeting DNA fragment, which is necessary to this technique. These various attempts are described in section 4.3.

In addition, before targeting could commence, the recombination mediated targeting technique implemented here required the introduction of a recombination plasmid into the DH10B *E. coli* cells that harbour pBAC14567. This plasmid carries genes that encode proteins that participate in, and drive homologous recombination, which is necessary for targeting to occur. A variety of recombination plasmids were tested and experimental success was finally achieved with pR6K $\alpha\beta\gamma$ (gift from Francis Stewart). Following the production of recombinant bacterial colonies it was necessary to analyse the BAC DNA from these clones to establish whether targeting had occurred. Upon preparation of the BAC DNA for this purpose, it was discovered that the abundance of recombination plasmid in the DNA preparation made this analysis impossible. Therefore the *E. coli* cells were first required to be cured of the recombination plasmid before restriction analysis, to establish targeting, could be performed. These points are discussed in detail in sections 4.4 and 4.7.2.

Targeting of the *lox511/kan/lox511/HSVtk* recombination cassette to the downstream position was conclusively established by restriction analysis, as described in section 4.5. Recombination mediated cloning was then applied to target the second recombination cassette 5' of HS-26, to the 'upstream' position (again see Fig. 2.1 for nomenclature). This second targeting event delivered a recombination cassette containing the *loxP/Hprt* Δ 5' cassette, as described by Smith and colleagues in 1995 (Smith et al., 1995; Van Deursen J, 1995), termed within this thesis as *loxP/3'hprt*. The isolation of clones targeted with this cassette was aided by the inclusion of a blasticidin (*bsd*) drug resistance gene as a selectable

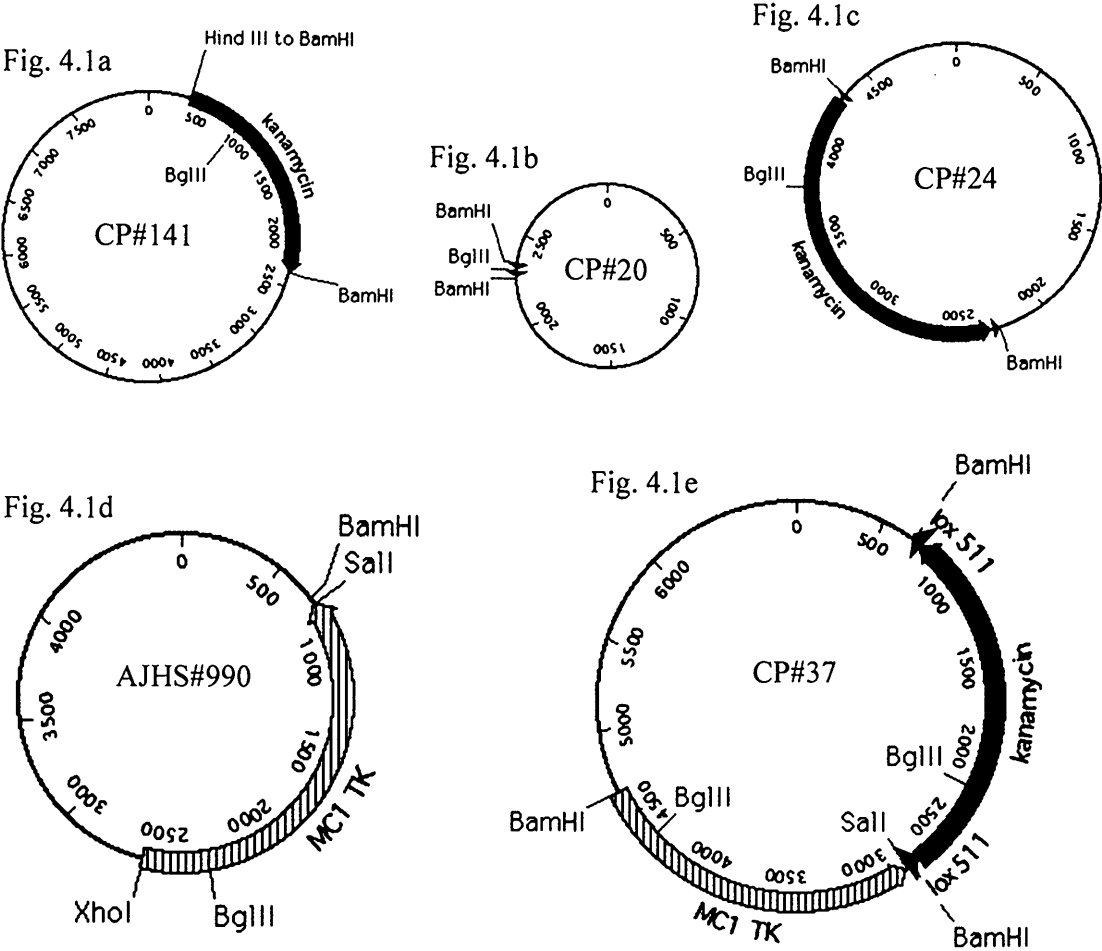
marker. This second recombination cassette also included an *f_{rt}* site, the purpose of which will be explained later, in Chapter 5. The design and construction of the *bsdloxP/3'hprt/I Sce I/f_{rt}* recombination cassette is described in section 4.6 and its subsequent targeting is described in section 4.7. Finally, experiments to confirm that the *lox* sites targeted to the BAC are functional are described in section 4.8.

4.2 Design and construction of the BAC downstream targeting construct

The alpha-globin downstream position in pBAC14567 was targeted with a *lox511/kan/lox511/HSVtk* recombination cassette. The purpose of including an *HSVtk* gene within this cassette was to use it as a negative marker in the selection strategy for the large interval RMCE event. The inclusion of the kanamycin gene was to facilitate the isolation of bacterial colonies harbouring targeted BACs. The purpose of *floxing* the kanamycin gene with *lox511* sites was to enable its excision and therefore its potential re-use as a selection marker for further modifications to the BAC, as desired. The *in vivo* or *in vitro* SSR between the *lox511* sites to excise the kanamycin gene would therefore leave a single *lox511* site, which would then be capable of participating in the large interval RMCE event. The construction of the *lox511/kan/lox511/HSVtk* recombination cassette is described in Fig. 4.1.

There were three important considerations concerning the targeting of the *lox511/kan/lox511/HSVtk* recombination cassette to the downstream position in the BAC. The first was that in order to affect negative selection the *HSVtk* gene needed to be excluded from the interval that would be exchanged during the large interval RMCE event. Thus, it was necessary to position the *HSVtk* gene outwith the region flanked by the hetero-specific *lox* sites, *loxP* and *lox511* (refer to Fig. 2.1). The second consideration was that the orientation of the *lox511* sites targeted to the BAC were required to be in the same orientation, relative to the alpha-globin sequences, as the *lox511* sites delivered to the CAT ES cell lines by the *lox511/hygromycin/I Sce I/lox511* targeting vector, CP#48 (see section 2.4). The third consideration was to target the *lox511/kan/lox511/HSVtk* recombination cassette to the identical sequence in the BAC as was targeted in the CAT ES cell lines with the *lox511/hygromycin/I Sce I/lox511* targeting vector. This was

Fig 4.1 Construction of the *lox511*/kan/*lox511*/HSV*tk* recombination cassette



The *Hind*III site in pKaX (CP#140) was first converted to a *Bam*HI site, by linearising the plasmid with *Hind*III, creating blunt ends with Klenow polymerase and dNTPs, and then ligation with the *Bam*HI linker NEB #1002 (see Fig. 4.1a). The kanamycin gene was then excised as a *Bam*HI fragment and cloned into the *Bgl*III site of CP#20 (shown in Fig. 4.1b and whose construction is described in Fig. 2.2). The resulting plasmid, CP#24, is shown in Fig. 4.1c. AJHS#990 (Fig. 4.1d) contains the sequence for the HSV thymidine kinase gene (HSV*tk*), denoted as 'TK' above. A *Bam*HI fragment, containing *lox511*/kanamycin/*lox511* sequences, was generated from CP#24 and cloned into the *Bam*HI site of AJHS#990. To confirm that the desired orientation of the *lox511* sites, in relation to the HSV*tk* gene, was obtained, the DNA from colonies resulting from this ligation were sequenced. A plasmid with the desired orientation of the *lox511*/kanamycin/*lox511* insert was obtained and the *Xho*I site 5' of HSV*tk* in this plasmid was converted to a *Bam*HI site to give plasmid CP#37, as depicted in Fig. 4.1e.

necessary in order to create a region that was flanked at identical genetic positions with hetero-specific *lox* sites in both the BAC and CAT ES cells, which would in turn permit the precise exchange of alpha-globin sequences during the large interval RMCE event.

4.3 Targeting the *lox511/kan/lox511/HSVtk* recombination cassette into the BAC

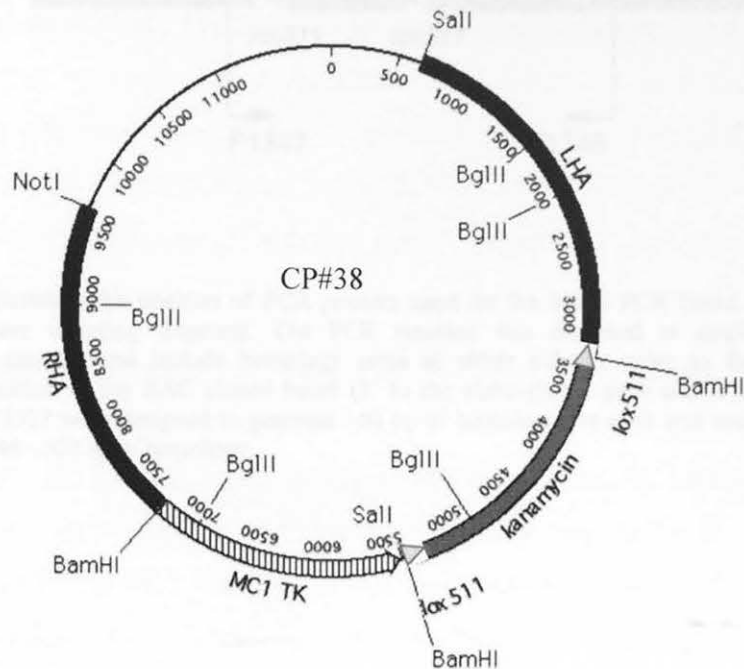
To produce a linear targeting DNA for recombination mediated targeting of *HSVtk* a number of different approaches were attempted before a successful strategy was found. The methods attempted are reported in this section.

4.3.1 PCR strategies for preparation of the targeting fragment

The first attempts to produce a linear targeting fragment were by PCR amplification. The PCR products were generated from the plasmid described in Fig. 4.2 as the template. It was reported that the efficiency of recombination mediated targeting increases in direct relation to the size of the homology arms (Zhang et al., 1998) and so the first pair of PCR primers, F1346 and F1347, were designed to produce homology arms of 300 bp either side of the *lox511/kan/lox511/HSVtk* recombination cassette (see Fig. 4.3). After several unsuccessful attempts to produce a PCR product a second set of primers, F53526 and F53527, were designed to yield 60 bp of homology either side of the recombination cassette. It was anticipated that the reduction in size of the PCR product would enable the PCR to be more efficiently optimised and reduce the chance of incorporating mutations into the homology arms, which could potentially be transferred undetected into the target site. The location of both primer sets is shown in Fig. 4.3.

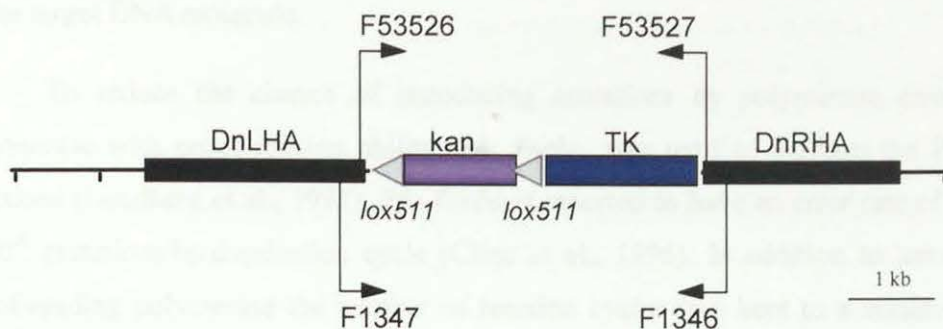
A significant consideration when using PCR to produce the double-stranded linear DNA for recombination mediated targeting is that it has the potential to introduce random mutations into the amplified template by polymerase error. The introduction of a mutation to a component of a targeting cassette could interfere with its ability to function. For example, a mutation in the *HSVtk* gene could have compromised the activity of the thymidine kinase gene and thus its effectiveness as

Fig 4.2 Construction of CP#38, the PCR template plasmid used to generate a linear targeting fragment of the *lox511*/kan/*lox511*/HSVtk recombination cassette



CP#37 (see Fig. 4.1e) was partially restricted with *Bam*HI and the 4 kb DNA fragment containing the *lox511*/kan/*lox511*/HSVtk sequences cloned into the *Bam*HI site in plasmid CP#45 (see Fig. 2.2e) to create CP#38. The orientation of the inserted sequences relative to the ES cell targeting homology arms was determined by *Bgl*II restriction analysis. The orientation of the *lox511* sites was then confirmed by sequence analysis with the oligonucleotides F1346 and F1347 (see Fig. 3.3) and determined to be consistent with the orientation of the *lox511* sites in ES cells targeted with CP#48.

Fig. 4.3 The position of PCR primers used to amplify the *lox511/kan/lox511/HSVtk* recombination cassette



The diagram illustrates the position of PCR primers used for the initial PCR based strategies to generate a linear targeting fragment. The PCR reaction was designed to amplify the TK recombination cassette and include homology arms at either side in order to target the downstream position in the BAC cloned insert (3' to the alpha-globin gene cluster). Primer pair F53526 and F53527 were designed to generate ~60 bp of homology for each arm and primer pair F1347 and F1346 ~300 bp of homology.

a negative selection marker. Alternatively, a mutation located within the homology arms of the targeting vector could prohibit homologous recombination, or at least reduce its efficiency and accuracy, and furthermore introduce unreported mutations to the target DNA molecule.

To reduce the chance of introducing mutations by polymerase error a polymerase with proof reading ability, *Pfu Turbo*, was used to perform the PCR reactions (Lundberg et al., 1991). *Pfu Turbo* is reported to have an error rate of 1.3×10^{-6} mutations/bp/duplication cycle (Cline et al., 1996). In addition to using a proof-reading polymerase the number of reaction cycles was kept to a minimum, between 15 and 25 cycles to reduce the probability of introducing mutations. This precaution made it difficult to optimise a PCR to generate enough product for subsequent electroporation.

Another disadvantage of using the PCR method was that the *Dpn* I treatment, used to digest the template plasmid from the PCR reaction contents (Zhang et al., 1998), was found to be inefficient. Therefore, unless the PCR products were gel purified, background colonies would arise from the electroporation of the PCR reaction, in which residual template plasmid persisted due to incomplete *Dpn* I restriction. The number of colonies arising from the template plasmid was always significant and made it necessary to replica plate large numbers of colonies to determine whether targeting had occurred (for the replica plating method see I.6.8 (f) in Appendix I). A template plasmid contamination in the electroporated DNA resulted in colony numbers such that a targeted event could not be distinguished even when replica plating 300 colonies or more.

4.3.2 Restriction digest strategies for the preparation of the targeting fragment

A significant disadvantage of the PCR strategies was the inability to produce adequate quantities of DNA for electroporation. To overcome this problem the PCR template plasmid CP#38, see Fig. 4.2, was implemented in a partial restriction strategy. The enzymes *Sac* I and *Kpn* I were used to liberate a linear targeting DNA with approximately 0.5 kb of homology either side of the *lox511/kan/lox511/HSVtk* recombination cassette. However, even when starting

with 150 µg of plasmid DNA, it was not possible to obtain sufficient quantities of DNA, following two rounds of partial digestion and subsequent gel purification, to successfully target pBAC14567.

4.3.2a The *Asc* I restriction strategy to generate the targeting fragment

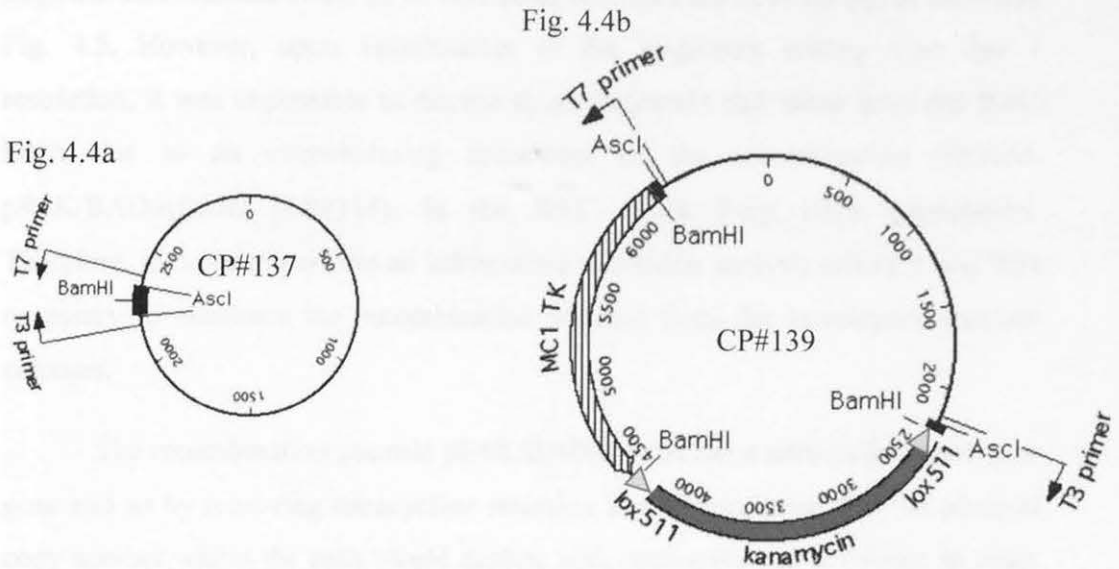
An alternative restriction-based strategy was therefore implemented, this involved building the homology arms into a plasmid by the insertion of a pair of annealed complementary oligonucleotides of 120 bp. These oligonucleotides were designed to incorporate a *Bam*HI site at their mid-point, used for the insertion of the *lox511/kan/lox511/HSVtk* cassette, and *Asc* I sites at their ends, which allowed the insertion of the oligonucleotides into a plasmid with a unique *Asc* I site. The strategy is summarized in Fig. 4.4. Although technically this approach was more tedious, it did have the advantage of eliminating the concerns previously discussed inherent in using PCR to generate the linear targeting fragment. The insertion of the *lox511/kan/lox511/HSVtk* cassette into the *Bam*HI site within the oligonucleotides created a plasmid that could then be *Asc* I restricted to liberate a linear targeting fragment, which included the recombination cassette bounded on each side by 60 bp of homology necessary to target the downstream position of the alpha-globin locus in pBAC14567.

4.4 Identifying targeted clones within a mixed population of kanamycin resistant colonies arising from the electroporation of the *lox511/kan/lox511/HSVtk* cassette into *E. coli* cells harbouring pBAC14567

The electroporation of the *Asc* I linear targeting fragment containing the *lox511/kan/lox511/HSVtk* recombination cassette into *E. coli* cells harbouring pBAC14567 generated kanamycin resistant colonies. These colonies required screening to identify those that harboured a targeted BAC molecule from those that arose from non-targeting events.

The kanamycin resistant colonies were screened by restriction analysis. DNA from 20 kanamycin resistant clones, chosen at random, was prepared

Fig 4.4 Construction of CP#139, the plasmid used to generate an *AscI* restriction fragment to target the *lox511/kan/lox511/HSVtk* cassette



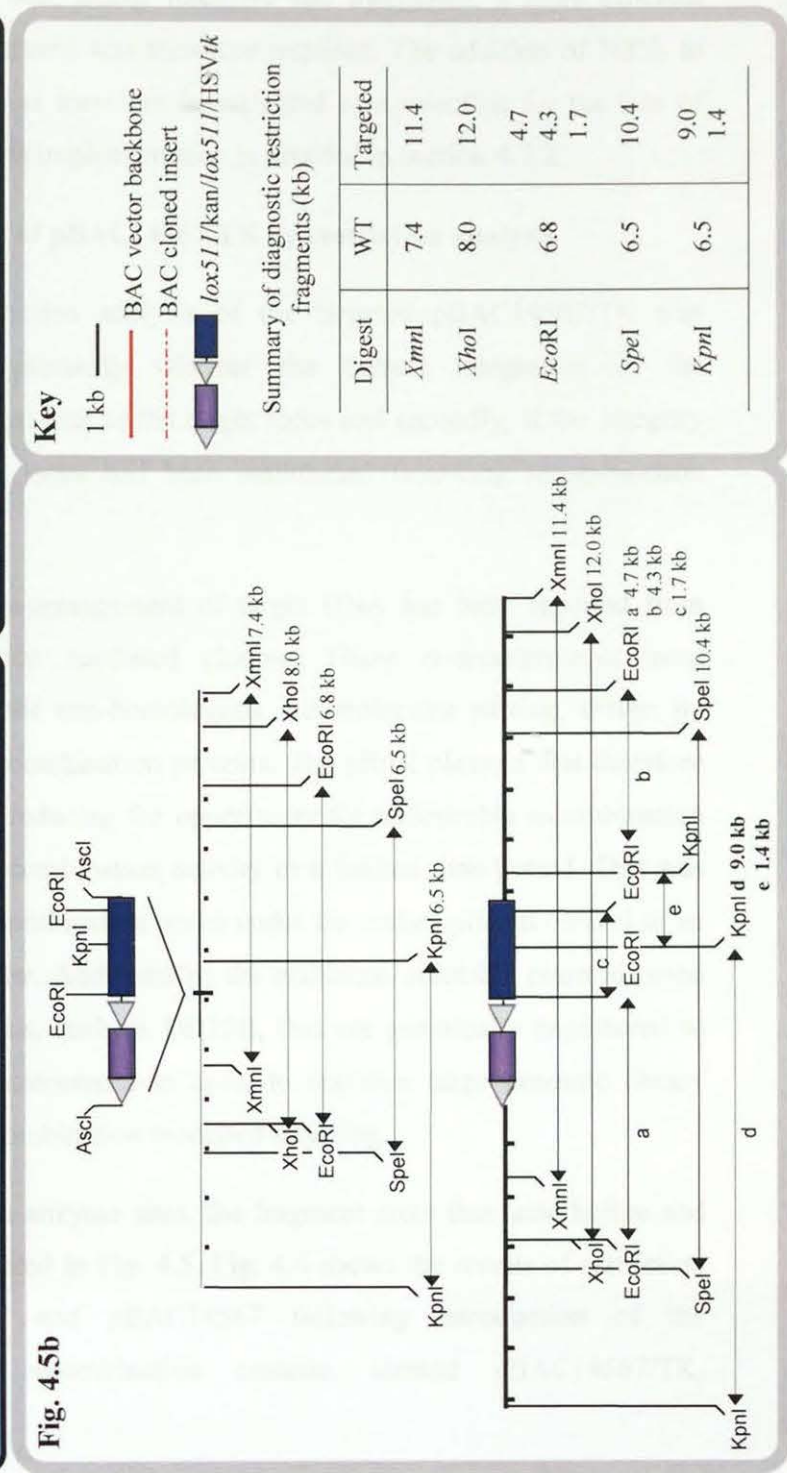
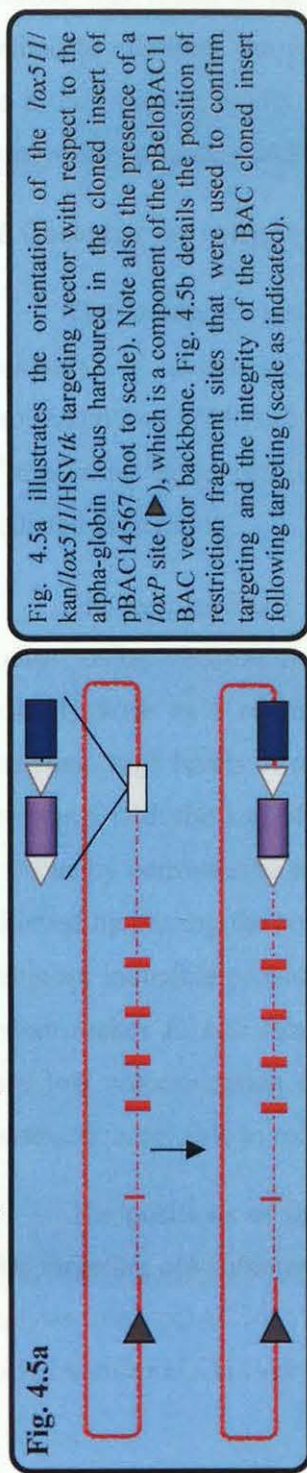
The complementary oligonucleotides F53524 and F53525 were designed with 120 bp of homology to the downstream targeting position of the alpha-globin locus within the BAC cloned insert. In addition a *Bam*HI site was included in their design at a position that was genetically identical to the *Bam*HI site between the ES cell targeting homology arms (as shown in Fig. 2.2e). Thus, both the ES cell and BAC recombination cassettes were delivered to genetically identical positions in ES cells and pBAC14567 respectively. F53524 and F53525 were annealed to create duplex DNA with *Asc*I compatible 3' single stranded ends. The annealed oligonucleotides were then ligated into the unique *Asc*I site of a plasmid, which had had an *Asc*I linker inserted between the *Kpn*I and *Sac*I sites of pBS(KS+) (gift from Melville Richardson) to give plasmid CP#137 (see Fig. 4.1). Insertion of the oligonucleotides was screened for by colony hybridization, using ³²P labelled F53525 as a probe. Positive clones were pooled, *Bam*HI restricted and treated with alkaline phosphatase. By partial *Bam*HI restriction a fragment containing the *lox511/kan/lox511/HSVtk* cassette was liberated from CP#37, this fragment was then ligated into the *Bam*HI site within the oligonucleotides in CP#137. Plasmid miniprep DNA was prepared from 40 kanamycin resistant colonies arising from the ligation and screened by *Bam*HI digest. 10 plasmids that gave *Bam*HI fragments indicating that the *lox511/kan/lox511/HSVtk* cassette had been inserted were sequenced with T3 and T7 primers, in order to determine the orientation of the cassette relative to the targeting homology arms. One of these plasmids, 37/Asc-21 or CP#139 (see Fig. 4.4b) was found to have the required orientation of the cassette and was restricted with *Asc*I to produce a 4.5 kb linear BAC targeting fragment.

separately from 5 ml overnight cultures by the BAC Quick Prep method (see I.6.1 (d)). The DNA was restricted with *Spe* I and the visualised results inspected for a fragment size increase of 6.5 kb to 10.4 kb as an indication of targeting, as shown in Fig. 4.5. However, upon visualization of the fragments arising from *Spe* I restriction, it was impossible to discern those fragments that arose from the BAC DNA due to an overwhelming abundance of the recombination plasmid, pR6K/BAD/ $\alpha\beta\gamma$ /*tet* (CP#114), in the BAC Quick Prep DNA preparation. Therefore, in order to perform an informative restriction analysis screen it was first necessary to eliminate the recombination plasmid from the kanamycin resistant colonies.

The recombination plasmid pR6K/BAD/ $\alpha\beta\gamma$ /*tet* has a tetracycline resistance gene and so by removing tetracycline selection it was anticipated that the plasmid copy number within the cells would deplete with successive cell divisions. In order therefore to obtain *E. coli* clones that had eliminated the recombination plasmid 40 kanamycin resistant colonies, selected at random, were re-streaked successively four times onto agar plates supplemented with kanamycin alone.

To then test whether the recombination plasmid had been eliminated the 40 clones were replica plated, first onto agar plates with both tetracycline and kanamycin selection, and secondly onto plates with kanamycin selection alone. Those clones that grew on the second plate but not the first were predicted to have eliminated the recombination plasmid. 6 of the 40 clones were tetracycline sensitive and therefore indicated that the recombination plasmid had been eliminated. These 6 clones were subsequently picked and grown overnight in a 5 ml culture for DNA preparation by the BAC Quick Prep method. The BAC DNA from these 6 clones was restricted with *Spe* I and upon visualisation it was conclusively established that all 6 had eliminated the recombination plasmid. Further, inspection of the *Spe* I fragment sizes determined that all 6 of the clones demonstrated the fragment size increase of 6.5 kb to 10.4 kb that indicated targeting (see Fig. 4.5). However, to conclusively determine whether targeting had occurred, and that no other unwanted re-arrangements of the BAC sequence had occurred, it was necessary to

Fig. 4.5 The restriction enzyme analysis used to determine and confirm *lox511/kan/lox511/HSVtk* targeting



scrutinise the BAC DNA from these clones, putatively assigned as targeted, by restriction analysis. This analysis is described in section 4.5 below.

The method of re-streaking colonies in order to eliminate the recombination plasmid was found to be both labour intensive and inefficient, a more efficient method to eliminate the plasmid was therefore required. The addition of NiCl₂ to the *E. coli* growth media was therefore investigated as a selection for the loss of tetracycline resistance and its implementation is detailed in section 4.7.2.

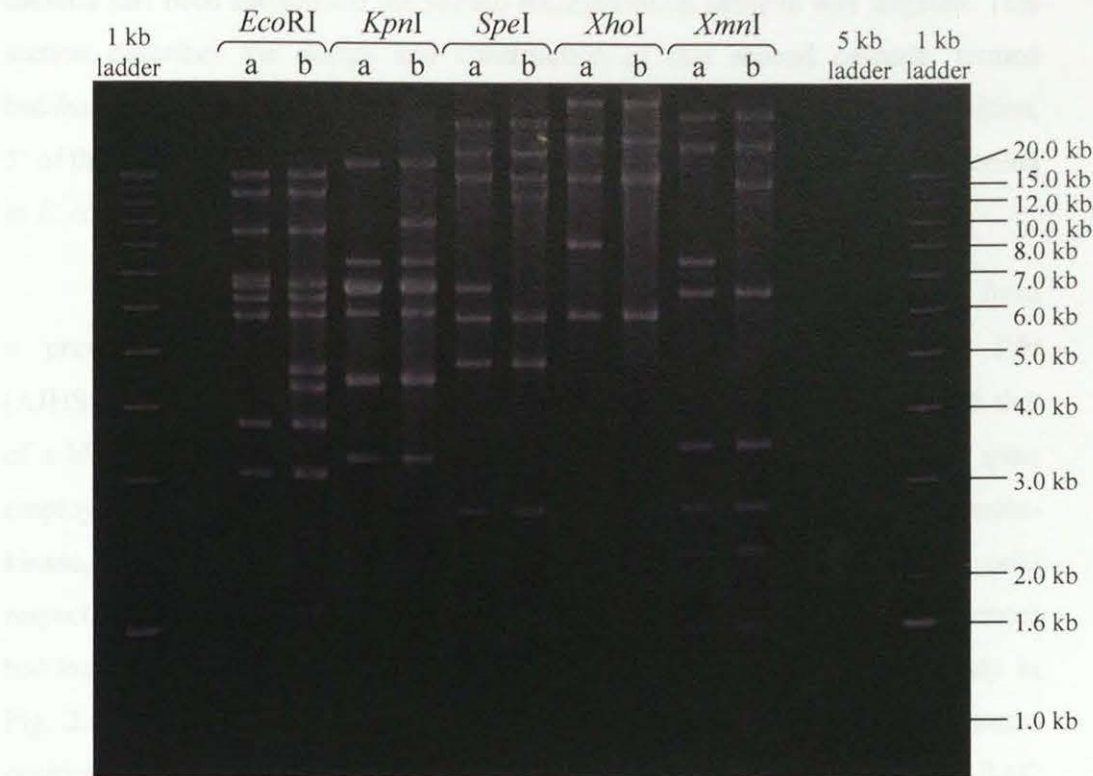
4.5 Checking the integrity of pBAC14567/TK by restriction analysis

The thorough restriction analysis of the targeted pBAC14567/TK was necessary to determine primarily whether the correct integration of the *lox511/kan/lox511/HSVtk* cassette to the target locus and secondly, if the integrity of the entire alpha-globin locus had been maintained following recombination mediated targeting.

The unintentional re-arrangement of target DNA has been reported from studies using recombination mediated cloning. These re-arrangements most probably arise as a result of non-homologous intramolecular pairing, driven by persistent, high levels of recombination proteins. The pR6K plasmid was therefore developed with the aim of reducing the opportunity for undesirable recombination to occur by constraining recombination activity to a limited time period. This was achieved by placing the recombination genes under the transcriptional control of an arabinose inducible promoter. Additionally, the arabinose inducible plasmid based system makes *E. coli* strains, such as DH10B, that are genetically engineered to have low recombination competence in order to maintain large genomic library constructs, amenable to recombination mediated targeting.

The positions of the enzyme sites, the fragment sizes that arise before and after targeting are summarized in Fig. 4.5. Fig. 4.6 shows the results of restriction analysis on pBAC14567 and pBAC14567 following introduction of the *lox511/kan/lox511/HSVtk* recombination cassette, termed pBAC14567/TK.

Fig 4.6 Restriction analysis of pBAC14567 targeted with the *lox511/kan/lox511/HSVtk* recombination cassette to give pBAC14567/TK



5 µg of DNA from each BAC, prepared using the Magnesil method, was restricted with *EcoRI*, *KpnI*, *SpeI*, *XhoI*, and *XmnI*. As shown: **a** indicates the untargeted BAC, pBAC14567 and **b** indicates pBAC14567 targeted with *lox511/kan/lox511/HSVtk*, to give pBAC14567/TK.

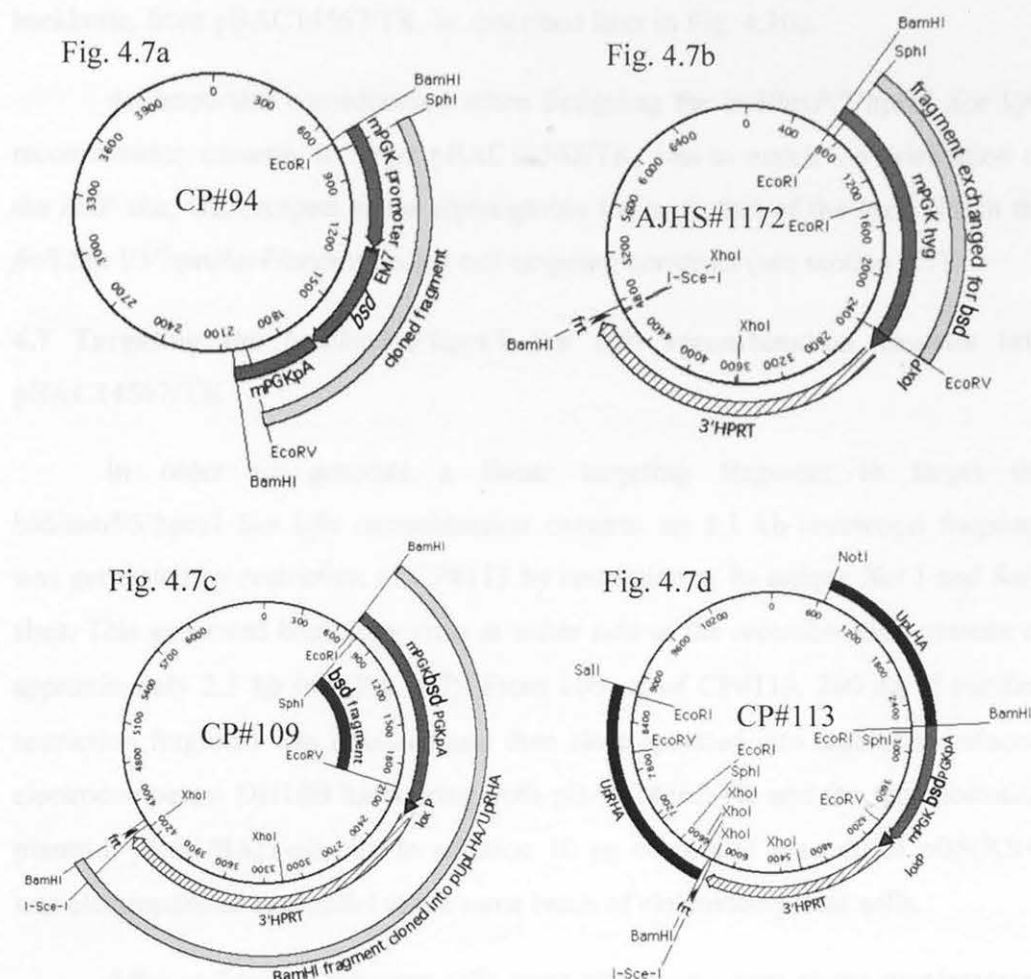
4.6 Design and construction of the BAC upstream targeting construct

Once the accurate delivery of the *lox511/kan/lox511/HSVtk* recombination cassette had been established the second recombination cassette was targeted. This section describes the design and construction of this second cassette, termed *bsd/loxP/3'hpvt/I Sce I/fvt*, and its targeted delivery to the upstream target position, 5' of the alpha-globin locus in pBAC14567/TK, by recombination mediated cloning in *E. coli*.

The *bsd/loxP/3'hpvt/I Sce I/fvt* recombination cassette was constructed from a previously existing plasmid, pBS/mPGKhyg/*loxP*/hpvt Δ 5'/I *Sce I/fvt* (AJHS#1172). The sequence encoding the hygromycin gene was replaced with that of a blasticidin resistant gene (*bsd*), as depicted in Fig. 4.7. The blasticidin gene employed has both prokaryotic (EM7) and eukaryotic (murine phospho-glycerate-kinase, mPGK) promoters, to enable its selection in *E. coli* and murine ES cells respectively. A *Bam*HI restriction fragment containing the sequences *bsd/loxP/3'hpvt/I Sce I/fvt*, was then inserted into CP#7 (described previously in Fig. 2.6c), which contains the homology arms employed to target the upstream position in ES cells (see Fig. 2.1). Thus, the *bsd/loxP/3'hpvt/I Sce I/fvt* BAC recombination cassette was positioned at the identical genetic location as the *fvt/I Sce I/5'hpvt/loxP/neomycin* ES cell recombination cassette, with respect to the alpha-globin locus. In addition, the orientation of the *bsd/loxP/3'hpvt/I Sce I/fvt* cassette within the ES cell homology arms was verified by restriction digest and sequencing. This was done to ensure that the orientation of the *loxP* site targeted to the BAC sequence would be consistent with the orientation of the *loxP* site previously targeted to the ES cell upstream position (see section 2.5).

The blasticidin gene of *bsd/loxP/3'hpvt/I Sce I/fvt* was not *floxed*, in contrast to the kanamycin gene in the construction of the *lox511/kan/lox511/HSVtk* cassette. The reason for this was that the *bsd* gene lies outwith the interval created by the hetero-specific *lox* sites in the BAC, which will be the interval exchanged in the desired large interval RMCE event, as described in Fig. 1.3. Therefore the *bsd* gene should not be transferred to the ES cell genome.

Fig. 4.7 Construction of CP#113, the *bsd/loxP/3'hprt/I Sce I/frt* recombination cassette and BAC targeting vector to target the upstream position of the alpha-globin locus



CP#94 (HN381), shown in Fig. 4.7a, was restricted with *Eco*RV and *Sph*I and the 1.3 kb blasticidin gene fragment (mPGKpEM7bsdP) gel purified, quantitated and ligated with the 5.4 kb *Eco*RV and *Sph*I fragment of AJHS#1172 (pPGKhyg/*loxP*/3'HPRT/*I Sce I*/*frt*), shown in Fig. 4.7b. The desired plasmid was screened for by *Bam*HI restriction, with the correct plasmid having fragments of 3.7 and 3.0 kb. The resulting plasmid was termed CP#109, as shown in Fig. 4.7c. The *bsd/loxP/3'hprt/I Sce I/frt* cassette, from CP#109, was then cloned into CP#7 (see Fig. 2.6c) to create a plasmid from which a linear targeting DNA could be produced, shown in Fig. 4.7d. This was achieved by restricting CP#109 with *Bam*HI and ligating the resulting 3.7 kb fragment into *Bam*HI restricted and alkaline phosphatase treated CP#7. Integration of *bsd/loxP/3'hprt/I Sce I/frt* cassette was screened for by *Bam*HI restriction and the desired orientation of the *bsd/loxP/3'hprt/I Sce I/frt* cassette was determined by restriction with *Eco*RI, *Eco*RV, *Xho*I. The linear targeting DNA, required for recombination mediated targeting, was generated from CP#113 by restriction with *Not*I and *Sal*I.

Although, upon inspection of the pBAC14567/TK sequence it was evident that the *loxP* site within the BAC backbone was in a configuration such that Cre mediated SSR would delete the *bsd* gene and a small region, 70 bp of BAC vector backbone, from pBAC14567/TK, as described later in Fig. 4.10c.

An important consideration when designing the *bsd/loxP/3'hprt/I Sce I/frt* recombination cassette, to target pBAC14567/TK, was to match the orientation of the *loxP* site, with respect to the alpha-globin locus, to that of the *loxP* site in the *frt/I Sce I/5'hprt/loxP/neomycin* ES cell targeting construct (see section 2.5).

4.7 Targeting the *bsd/loxP/3'hprt/I Sce I/frt* recombination cassette into pBAC14567/TK

In order to generate a linear targeting fragment to target the *bsd/loxP/3'hprt/I Sce I/frt* recombination cassette, an 8.1 kb restriction fragment was generated by restriction of CP#113 by restriction at its unique *Not I* and *Sal I* sites. This generated homology arms at either side of the recombination cassette of approximately 2.5 kb (see Fig. 4.7). From 100 µg of CP#113, 260 ng of purified restriction fragment was obtained and then electroporated into arabinose induced, electrocompetent DH10B harbouring both pBAC14567/TK and the recombination plasmid, pR6K/BAD/ $\alpha\beta\gamma$ /*tet*. In addition 10 pg of control plasmid of pBS(KS+) was electroporated in parallel to the same batch of electrocompetent cells.

After an hour of recovery cells were plated onto agar plates supplemented with blasticidin selection. Cells with a competency of 5×10^6 / pg DNA, as determined by ampicillin resistance, produced 171 blasticidin resistant colonies. Clones arising after 24 hours of incubation at 37°C were picked and replica plated onto agar plates with different antibiotic selections added, in order to isolate clones that harboured the desired targeting event.

4.7.1 Screening colonies arising from electroporation of the *bsd/loxP/3'hprt/I Sce I/frt* recombination cassette

As discussed previously, following targeting, the resulting clones were required to be cured of the tetracycline resistance recombination plasmid,

pR6K/BAD/ $\alpha\beta\gamma$ /*tet*, in order to obtain a BAC DNA preparation that could be analysed by restriction analysis, to identify targeted clones.

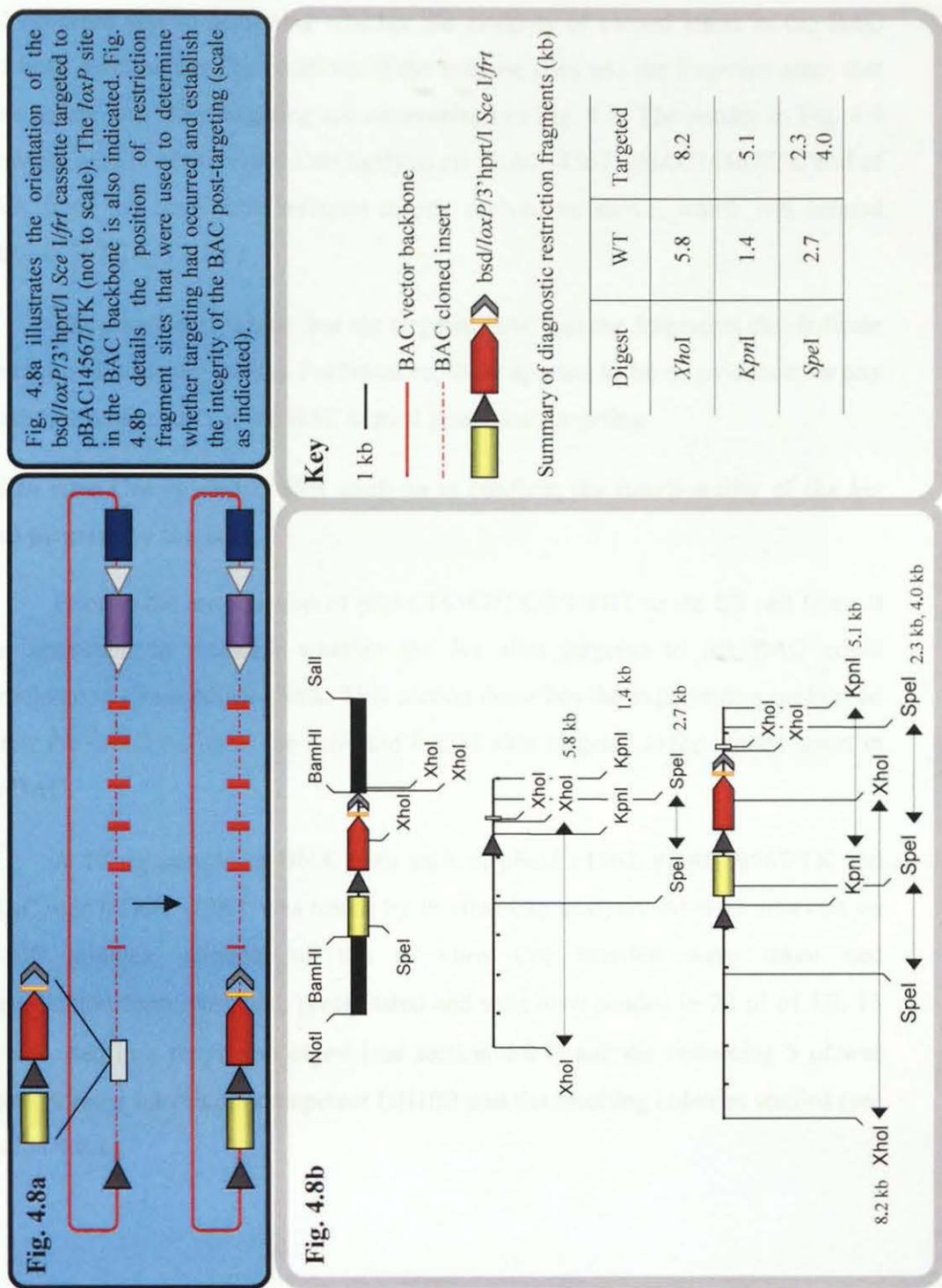
4.7.2 Elimination of the recombination plasmid by culture with NiCl₂

The primary screen to isolate a targeted integration of the recombination cassette was by restriction digest. As previously discussed, restriction analysis necessitated the prior elimination of the pR6K/BAD/ $\alpha\beta\gamma$ /*tet* recombination plasmid (see section 4.4). To achieve this rapidly and effectively for hundreds of blasticidin resistant colonies required a better procedure to select for the loss of the tetracycline resistance marker that is carried on pR6K/BAD/ $\alpha\beta\gamma$ /*tet*. This was achieved by supplementing the agar and culture broth, in which the colonies were grown, with NiCl₂ (see I.6.8(e)). This method has been reported to select for the loss of tetracycline resistance in *E. coli* more effectively than fusaric acid based selection (Podolsky et al., 1996).

20 of the 171 (see section 4.7 above) blasticidin resistant colonies were picked and cultured in 5 ml of broth overnight supplemented only with blasticidin. The following day 1 ml of the culture was used to inoculate broth supplemented with blasticidin and 5mM NiCl₂. The culture was grown overnight and a 100 μ l plated onto agar plates supplemented with blasticidin alone. The colonies that arose were replica plated, first onto agar plates supplemented with blasticidin and tetracycline and secondly to plates supplemented with blasticidin and 5mM NiCl₂.

DNA from each of the colonies, that survived on the blasticidin and NiCl₂ supplemented plates and that had therefore eliminated the recombination plasmid, was prepared by the BAC 'Quick Prep.' method. The DNA was then restricted by *Spe* I to identify targeting of the *bsd/loxP/3' hprt/I Sce I/frt* construct (see Fig. 4.8). The presence of a 4 kb fragment was taken to indicate targeting and this was evident from 11 of the 20 clones tested. All 20 clones demonstrated the absence of the recombination plasmid. A single clone was then chosen to perform more extensive restriction analysis, as described in the following section.

Fig. 4.8 Restriction digest screening strategy employed to determine *bsd/loxP/3' hprt/1 Sce I/frrt* targeting



4.7.3 Checking the integrity of pBAC14567/TK/3'HPRT by restriction analysis

The DNA from a single clone, that appeared to be targeted according to *Spe* I restriction analysis described above, was prepared by the Magnesil method. The DNA was restricted with *Sma* I, *Spe* I and *Xho* I, in order to confirm that targeting had occurred and to determine whether the integrity of cloned insert in the BAC had been maintained. The positions of the enzyme sites and the fragment sizes that arise before and after targeting are summarized in Fig. 4.8. The results in Fig. 4.9 show the results of this restriction analysis on pBAC14567, pBAC14567/TK and of DNA from the blasticidin resistant colony derived as above, which was termed pBAC14567/TK/3'HPRT.

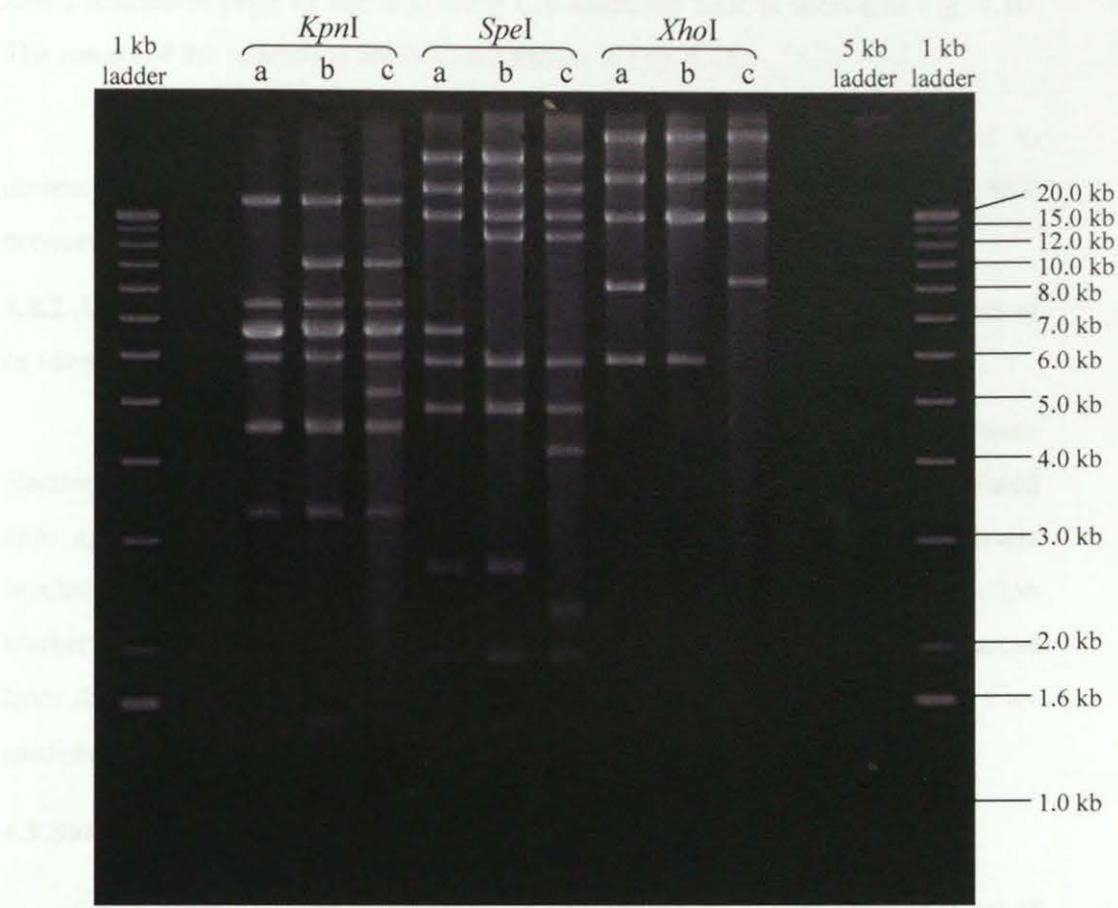
The data clearly show that the targeted BAC has the fragments that indicate the required targeting events. Furthermore, there appears to be no evidence for any gross re-arrangements of the BAC cloned insert post-targeting.

4.8 *In vitro* Cre-mediated SSR analysis to confirm the functionality of the *lox* sites targeted to the BAC

Prior to the introduction of pBAC14567/TK/3'HPRT to the ES cell lines, it was important to establish whether the *lox* sites targeted to the BAC could participate in Cre-mediated SSR. This section describes the experiments performed to test the functionality of the *loxP* and *lox511* sites targeted to the cloned insert in the BAC.

A 10 µg sample of DNA from each of pBAC14567, pBAC14567/TK and pBAC14567/TK/3'HPRT was tested by *in vitro* Cre analysis. At time intervals up to 30 minutes, aliquots of the *in vitro* Cre reaction were taken and phenol/chloroform extracted, precipitated and then resuspended in 20 µl of TE. 15 µl was used in a restriction digest (see section 4.8.1) and the remaining 5 µl was electroporated into electrocompetent DH10B and the resulting colonies studied (see section 4.8.2).

Fig. 4.9 Restriction analysis of pBAC14567/TK targeted with the *bsd/loxP/3'hprt/I Sce I/frt* recombination cassette



5 μ g of DNA from each BAC, prepared using the Magnesil method, was restricted with *KpnI*, *SpeI*, and *XhoI*, as shown: **a** indicates the untargeted BAC, pBAC14567; **b** indicates pBAC14567/TK and **c** indicates pBAC14567/TK targeted with *bsd/loxP/3'hprt/I Sce I/frt*, to give pBAC14567/TK/3'HPRT.

4.8.1 *Xho* I restriction analysis on *in vitro* Cre reactions

The 15 µl aliquots of BAC DNA that had been used in the *in vitro* Cre reaction were restricted with *Xho* I. An outline of the predicted fragment sizes from *Xho* I restriction prior to and following Cre-mediated SSR is shown in Fig. 4.10. The results of the restriction analysis are shown in Fig. 4.11.

The colonies arising from the electroporation were replica plated to determine which selection markers had been deleted as a consequence of SSR between homologous *lox* sites.

4.8.2 Analysis of colonies arising from electroporation of the DNA products of *in vitro* Cre-mediated SSR

As mentioned above, 5 µl aliquots of the *in vitro* Cre reaction were electroporated into DH10B cells. The cells were allowed to recover and then plated onto agar supplemented with chloramphenicol. Colonies arising after 24 hours incubation at 37 °C were then replica plated to determine which of the selection markers had been deleted. The results of the replica plating from colonies that arose from the 5 µl aliquot from the pBAC14567/TK/3'HPRT molecule, following Cre-mediated SSR are summarised in Fig. 4.12.

4.9 Summary

This chapter describes the genetic modifications made to the cloned insert of pBAC14567 in order to introduce the *lox511/kan/lox511/HSVtk* and then the *bsd/loxP/3'HPRT* targeting constructs. These sequences were targeted to positions that result in flanking the entire alpha-globin locus with *loxP* and *lox511* in preparation to test the RMCE strategy, as outlined in section 1.5.4(a).

These modifications were achieved using the relatively new technology of recombination mediated cloning. From the experience of implementing this technology a number of important generalisations can be made concerning the practical applications of this strategy of genetic manipulation.

Fig. 4.10 The analysis performed to test the functionality of the *lox* sites targeted to pBAC14567

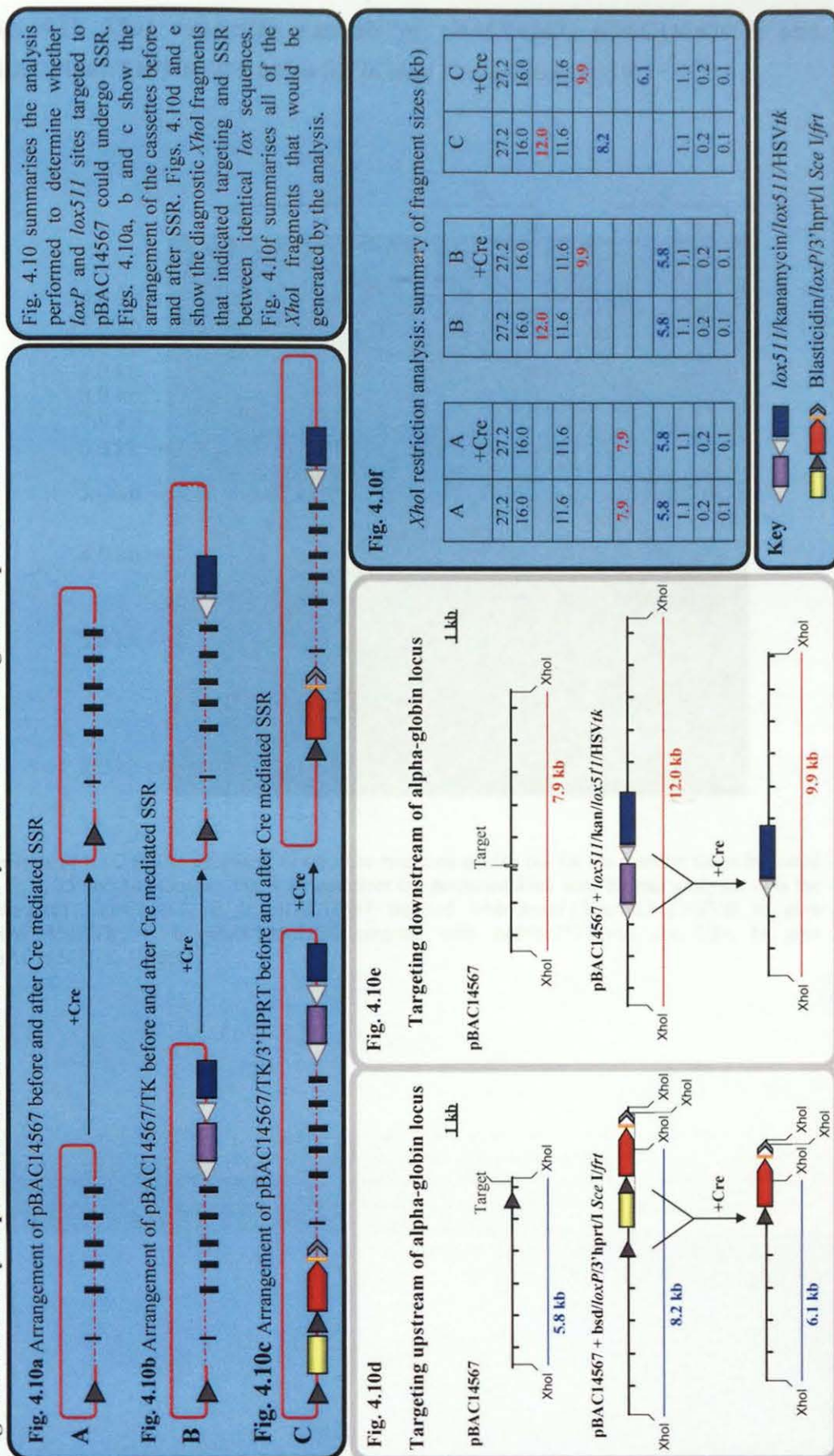
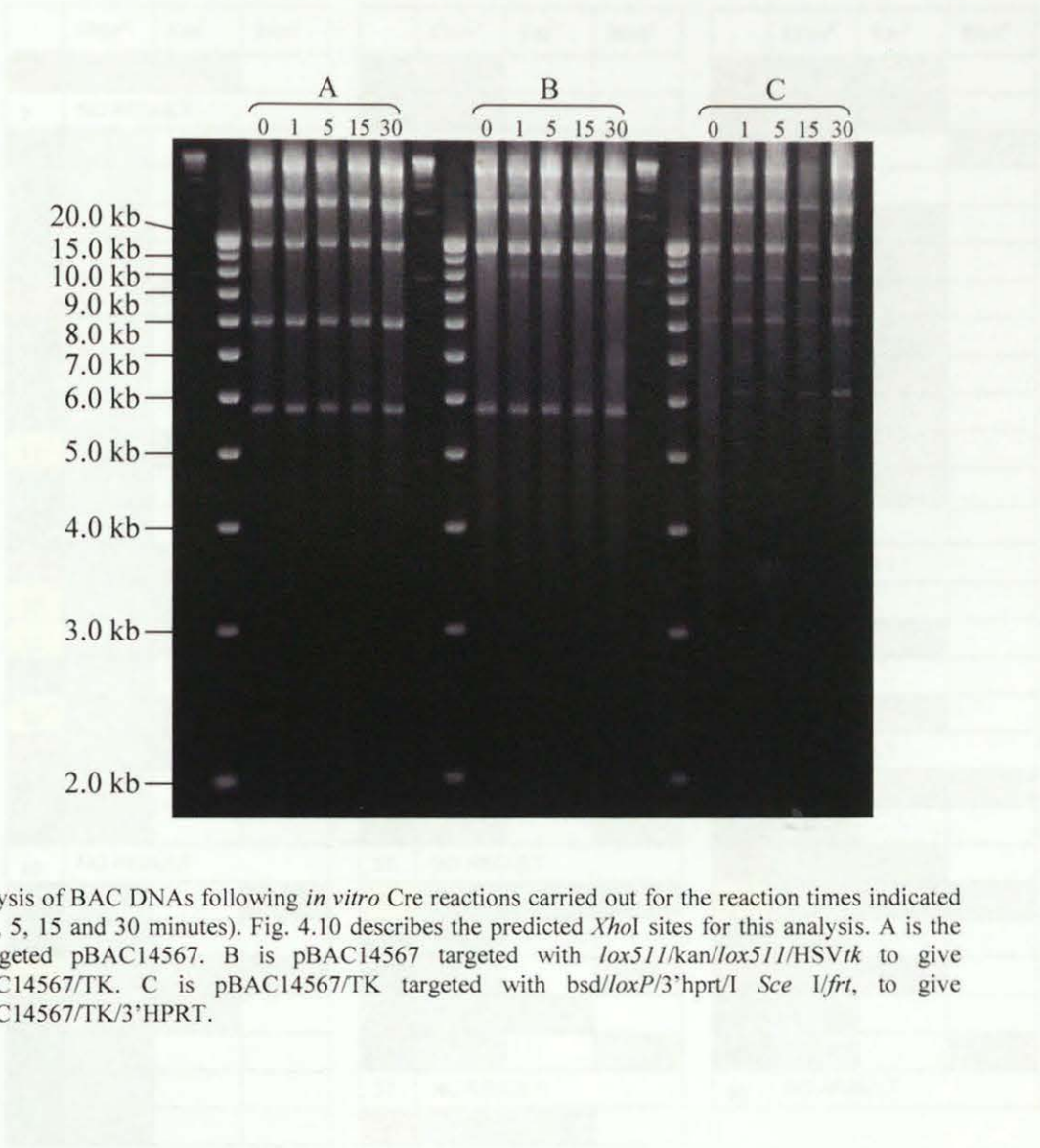


Fig. 4.11 *XhoI* restriction analysis of pBAC14567, pBAC14567/TK and pBAC14567/TK/3'HPRT following *in vitro* Cre-mediated SSR



Analysis of BAC DNAs following *in vitro* Cre reactions carried out for the reaction times indicated (0, 1, 5, 15 and 30 minutes). Fig. 4.10 describes the predicted *XhoI* sites for this analysis. A is the untargeted pBAC14567. B is pBAC14567 targeted with *lox511/kan/lox511/HSVtk* to give pBAC14567/TK. C is pBAC14567/TK targeted with *bsd/loxP/3'hprt/l Sce I/frt*, to give pBAC14567/TK/3'HPRT.

Fig. 4.12 The results of replica plating chloramphenicol resistant colonies arising from *in vitro* Cre-mediated SSR on pBAC14567/TK/3'HPRT

	Chlor ^R	Kan ^R	Blast ^R		Chlor ^R	Kan ^R	Blast ^R		Chlor ^R	Kan ^R	Blast ^R
1				30				59			
2	NO RESULT			31				60			
3				32	NO RESULT			61			
4				33				62			
5				34				63			
6				35				64			
7				36				65			
8				37				66			
9				38				67			
10				39				68			
11				40				69			
12				41				70			
13				42				71			
14				43				72	NO RESULT		
15				44				73			
16				45				74			
17				46				75			
18				47				76			
19				48				77	NO RESULT		
20				49				78			
21				50				79			
22	NO RESULT			51	NO RESULT			80			
23				52				81			
24				53				82			
25				54				83			
26				55				84			
27				56				85			
28				57	NO RESULT			86	NO RESULT		
29				58							

The table summarises the resistance profiles of the *E. coli* cells following Cre-mediated SSR. Key: Chlor^R - Chloramphenicol resistance; Kan^R - Kanamycin resistance; Blast^R - Blasticidin resistance. The grey shading indicates resistance and each clone is grouped by its resistance profile. 14% of the clones were kan^R/bsd^R and were therefore unchanged following SSR (yellow). 17% were kan^S/bsd^R and had therefore deleted the *floxed* kanamycin selection marker alone (blue). 33% were kan^R/bsd^S and had therefore deleted the *floxed* blasticidin selection marker alone (red). 36% were kan^S/bsd^S and had therefore deleted both the kanamycin and blasticidin selection markers (green).

Firstly, the PCR strategies tested to produce a linear targeting DNA were found to be inefficient to produce adequate quantities of DNA that could drive the efficiency of the recombination mediated cloning technique and achieve BAC targeting. For this reason restriction based strategies were employed to produce the linear targeting DNA for both the recombination cassettes. Although this involved using additional cloning steps to generate the linear targeting DNA, this approach at least eliminated some of the other concerns previously discussed concerning the use of PCR.

Secondly, it was necessary to conduct a primary screen of colonies arising from electroporation to distinguish targeted colonies from background non-targeted colonies. Most of these background colonies were found to arise from the contamination of the targeting fragment, by the plasmid the fragment had been derived from. In the case of generating the targeting fragment by PCR, the incomplete digestion of the template plasmid by *Dpn* I caused its persistence. In the case of generating the targeting fragment by restriction, the incomplete digestion and subsequent inefficient gel purification of the template plasmid caused the contamination of the targeting fragment.

A primary screen was therefore employed to distinguish those colonies that harboured a BAC in which the linear targeting fragment had targeted, from those in which harboured the intact template plasmid, as a result of electroporating a contaminated linear targeting DNA.

Thirdly, although a primary screen could have been conducted by PCR, the definitive analysis of a targeted event required restriction of the targeted BAC DNA. However, in order to perform thorough restriction analysis on targeted colonies the recombination plasmid needed to be eliminated from the *E. coli* cells. Since the recombination plasmid exists within the *E. coli* cell at a high copy number, and the BAC as a single copy, preparation of the BAC DNA by the 'Quick Prep.' method resulted in a large amount of recombination plasmid DNA included in the total DNA preparation. Therefore upon restriction analysis, the bands arising

from BAC DNA restriction were obscured from view by those arising from the recombination plasmid.

Thus, although in principle the recombination mediated cloning technique appears straightforward, in practise it was found to be quite tedious. However, on a positive note, the disadvantages associated with recombination mediated cloning, of undesirable re-arrangements to the target molecule during targeting, was not encountered to any great extent here.

Once the BAC was successfully targeted with both recombination cassettes it was necessary to determine whether the *lox* sites could participate in Cre-mediated SSR. The *in vitro* Cre experiment demonstrates that the *floxed* selection markers could be deleted from the BAC by Cre-mediated SSR and therefore retained functionality. In addition, the deletion of the markers with the *in vitro* reaction does not go to completion but reaches a state of equilibrium. This results in generating a mixed population of BAC molecules that harbour either excised or non-excised versions of either of the bacterial selection markers (see Fig. 4.11). It was therefore possible to isolate, by electroporation of Cre recombinase treated pBAC14567/TK/3'HPRT, a variety of recombinant BACs that had deleted neither, either or both of the selection markers, illustrating further the potential ways in which BACs may be variously manipulated (see below).

The other motivation to performing the *in vitro* Cre analysis on the targeted BAC was to observe the tendency for the *loxP* and *lox511* sites to participate in undesirable SSR. No evidence was generated to suggest that the *loxP* and *lox511* sites underwent SSR with each other. It was therefore predicted that any undesirable SSR that may be taking place, was occurring at such a frequency as to be out-competed by the desired SSR reactions between identical *lox* sites. This was encouraging for the success of the proposed large interval RMCE experiment, since it suggested that the *lox* sites in the BAC would preferentially interact with their identical sites in the ES cell genome, thereby generating the desired large interval RMCE event.

The ability to identify and recover recombination intermediates from an *in vitro* Cre reaction by replica plating illustrates the potential by which the BAC could be manipulated. Specifically with regard to the alpha-globin locus it would be possible to *flox* each of the structural genes sequentially, reporting each targeting event with a different antibiotic selection marker, which would be included within the *floxed* region. Upon Cre-mediated deletion of the *floxed* gene the recombination events could be determined by replica plating to report the loss of the selection marker. Using this method it would be possible to rapidly isolate BACs with any combination of deleted and re-arranged structural genes. Currently it is not possible to rapidly generate deletion constructs by such a directed method. The ability to then transfer these mutations to an ES cell *in vitro* system for functional analysis would further expediate the study of the alpha-globin locus. For example, the mutated ES cell lines could be differentiated along an erythroid lineage (Baird et al., 2001) and the effect of mutation on the role of the structural genes during erythroid specification and blood island formation determined.

Chapter 5

LARGE INTERVAL RECOMBINASE MEDIATED CASSETTE EXCHANGE

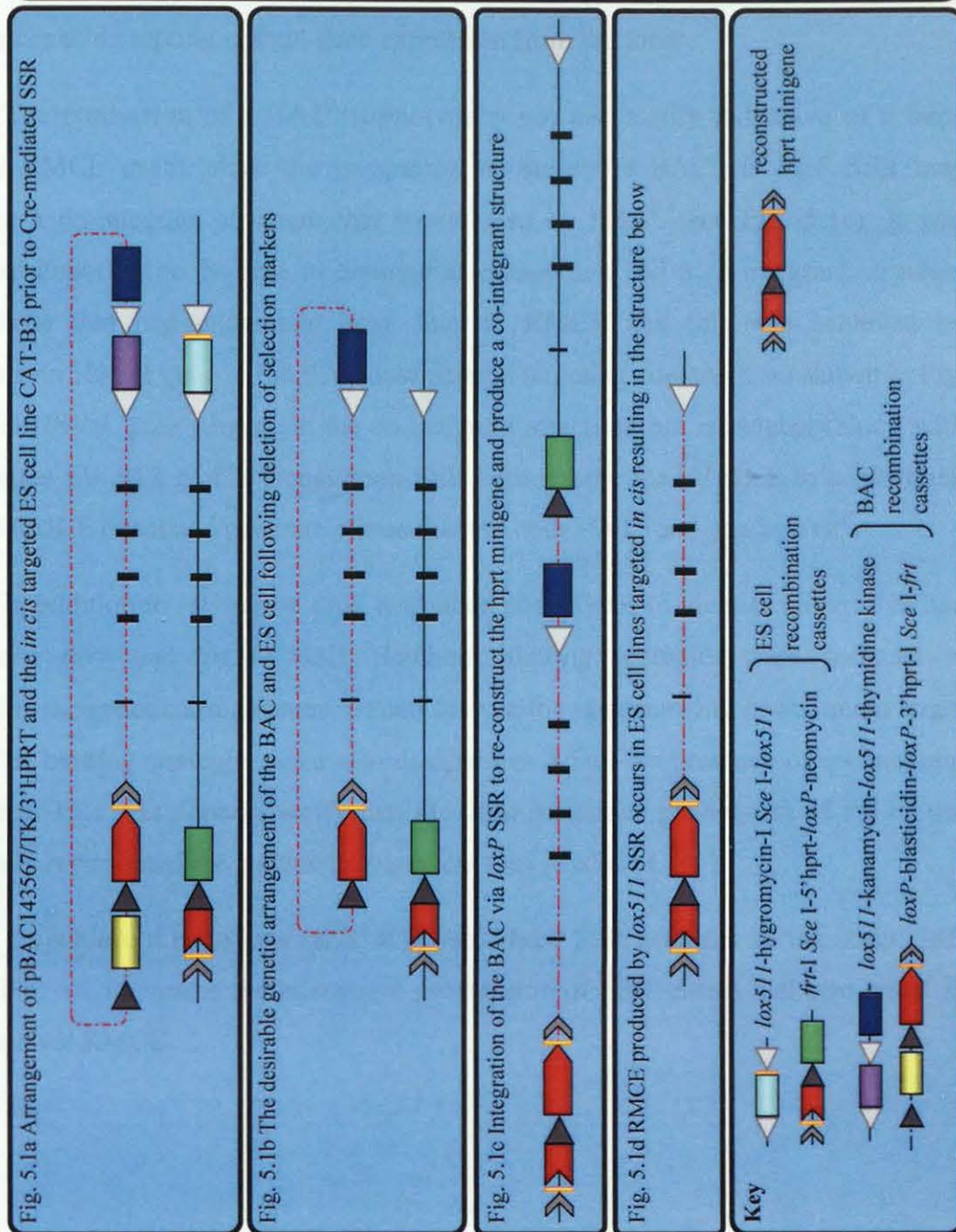
5.1 Introduction

Thus far, this thesis has described the production of the components necessary to accomplish a large interval RMCE event. Chapter 2 described the production of double targeted ES cell lines (CAT lines) and Chapter 3 described the experiments that were conducted to test whether the *lox* sites targeted to the CAT lines were functional and in a desired arrangement. The production of a targeted BAC and the functional assessment of the *lox* sites targeted to its cloned insert, were described in Chapter 4.

This chapter now describes the results of experiments that brought together the CAT ES cell lines and the targeted BAC, in order to achieve large interval RMCE. The experiments aimed to mediate a RMCE event of 64 kb, representing the exchange of the entire murine alpha-globin locus in ES cells with its equivalent locus from a BAC.

Fig. 5.1 illustrates the arrangement of all of the components of the large interval RMCE strategy. Additionally, Fig. 5.1 illustrates one of the possible combinations of SSR events that would result in the generation of the desired RMCE event.

Fig. 5.1 Replacement of the ES cell locus with the locus residing in the BAC: large interval RMCE



These figures depict just one possible sequence of SSR events that would result in a RMCE event between the CAT-B3 *in cis* line and pBAC143567/TK/3'HPRT.

The production of a co-integrant structure via *loxP* SSR is shown since this event can be selected for using HAT supplemented media and may be isolated for clones generated from both *in cis* and *in trans* lines. The resolution of the co-integrant structure from *in cis* generated clones is depicted in Fig. 5.1d.

Resolution by SSR between *lox511* sites to produce a RMCE event is predicted to be possible only for *in cis* targeted lines, since the *lox511* sites reside on the same chromosome following the production of the co-integrant.

To screen HAT^R clones, a probe internal to the exchange region was used, UpRP, see Fig 5.4. This probe should identify both the reconstructed *hprt* minigene and the co-integrant structure.

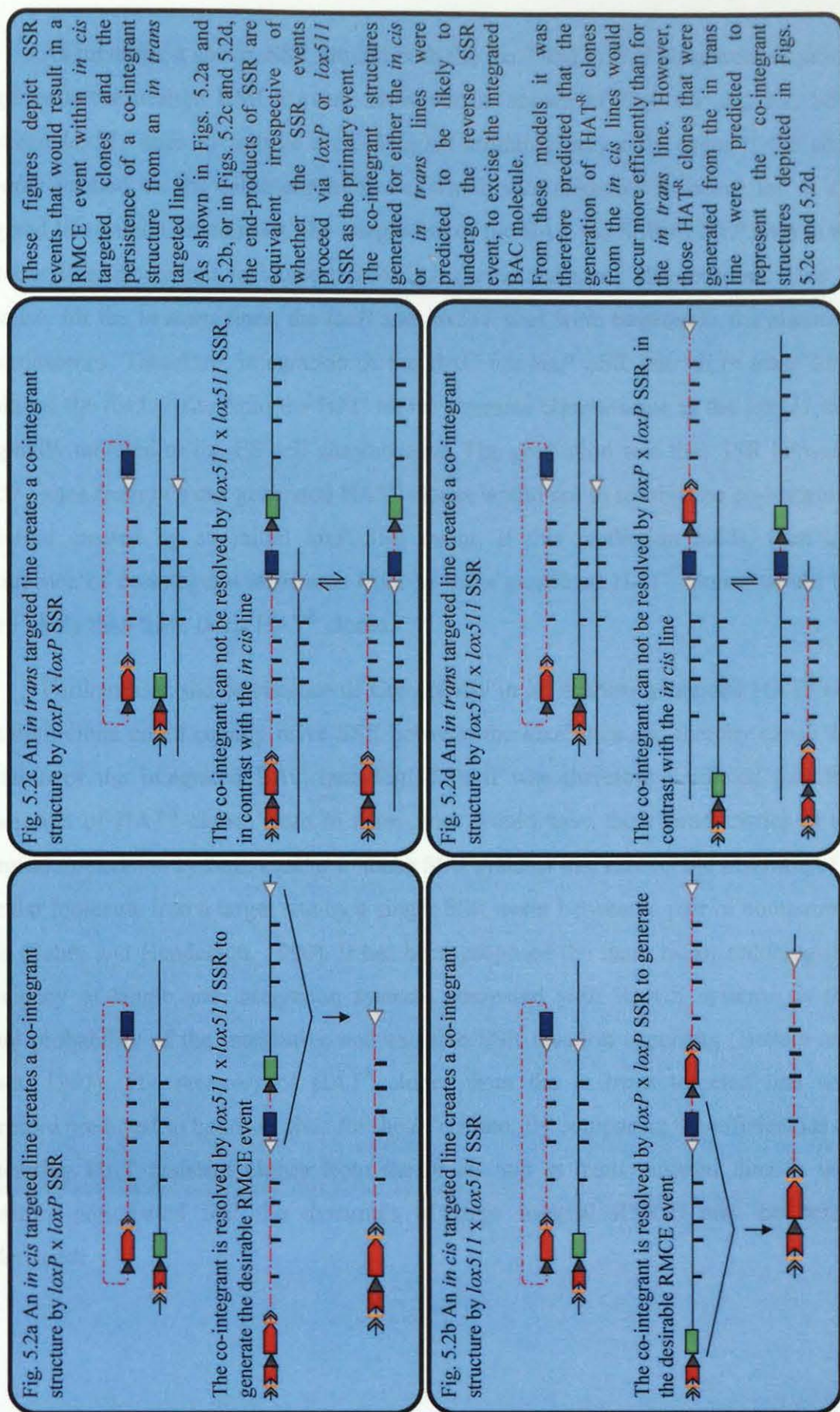
Fig. 5.1d illustrates the reconstruction of the *hprt* minigene around the *loxP* site in the ES cell genome, concomitant with the generation of the desired RMCE event. As illustrated, the reconstructed *hprt* minigene is flanked on either side by *frt* sites, which enables its deletion by the transient expression of FLP recombinase in cells with this genetic arrangement. The *frt* sites were included to incorporate flexibility into the strategy by enabling HAT selection to be implemented, should it be required, for subsequent targeting events. Additionally the removal of the *hprt* minigene would leave just a single *frt* and *lox511* site flanking the exchanged interval, thereby reducing the chances of disrupting normal gene expression from the locus.

The production of a HAT^R phenotype is not necessarily indicative of a large interval RMCE event, since the integration of the entire BAC via *loxP* SSR may generate a co-integrant structure that would also be HAT^R (see Fig. 5.1c). It was therefore important to be able to distinguish clones that had a co-integrant structure from those that had undergone large interval RMCE and this was achieved by including an *HSVtk* gene in the BAC downstream targeting construct. As shown in Fig. 5.1c, the *HSVtk* gene persists in the co-integrant structure, but is deleted (along with the original ES cells globin locus) upon SSR between the *lox511* sites, to achieve the desired RMCE event and generate clones that are both HAT^R and ganciclovir^R.

In addition to relying on drug resistance phenotypes to identify clones that had undergone *loxP* and *lox511* SSR, Southern blotting strategies were designed to determine the genetic arrangement around each of the upstream and downstream target sites. The blotting strategies were also designed to detect the presence of co-integrant structures. This was done to verify that the drug resistance phenotypes of the clones accurately represented the genetic arrangement they predicted.

To achieve a large interval RMCE event *loxP* SSR need not be the initial SSR event. Fig. 5.2 illustrates the alternative progression of SSR events that also result in large interval RMCE.

Fig. 5.2 SSR events to generate either a RMCE event or a co-integrant structure



As mentioned above, SSR around both the *loxP* and *lox511* sites were required to generate the desired RMCE event. However, as shown in Figs. 5.1 and 5.2, SSR between *lox511* sites to resolve a co-integrant structure may only occur if the sites become located on the same chromosome, which was predicted to occur for *in cis* targeted lines alone. In contrast, the integration of the BAC via SSR of *loxP* sites to an *in trans* line should not to locate the *lox511* site on the same chromosome. This is because, for the *in trans* lines, the *loxP* and *lox511* sites were targeted to the alternate chromosomes. Therefore, integration of the BAC via *loxP* SSR into an *in trans* line, positions the *lox511* site from the BAC on the alternate chromosome to the *lox511* site originally targeted to the ES cell chromosome. The prediction was that SSR between *lox511* sites from *in trans* generated HAT^R clones would not to resolve the co-integrant structure created by an initial *loxP* SSR event. If this prediction holds, then the occurrence of co-integrant structures from *in trans* generated HAT^R clones would be more likely than from *in cis* HAT^R clones.

Furthermore, the persistence of Cre activity in an *in trans* generated HAT^R co-integrant clone could equally drive SSR between the *loxP* sites and thereby cause the excision of the integrated BAC (see Fig. 5.2). It was therefore predicted that the generation of HAT^R clones from *in trans* lines would have the characteristics of an integration-excision system, similar to those SSR systems that rely on the insertion of a circular molecule into a target site by a single SSR event between a pair of compatible sites (Sauer and Henderson, 1990). It has been proposed the main factor reducing the efficiency of single site integration systems compared with RMCE systems, is the equal probability of the integrative and excisive SSR reaction occurring (Bethke and Sauer, 1997). The recovery of HAT^R clones from the *in trans* targeted line was therefore predicted to be lower than for the *in cis* line. By comparing the efficiencies of generating HAT resistant clones from the *in cis* and *in trans* targeted lines it was therefore anticipated that the dynamics of large interval RMCE may be better understood.

The targeting of ES cells to generate lines that could participate in the large interval RMCE event generated lines that had targeted either *in cis* or *in trans* (see Chapter 3). In order to test the strategy for large interval RMCE event the CAT-B3 line, which was assigned as targeted *in cis*, was selected. In parallel, it was decided to perform the equivalent experiment in the CAT-C10 line, which had been putatively assigned as targeted *in trans*. Although the CAT-C10 clone had generated data which conflicted with its *in trans* assignment (see Chapter 3), it was still chosen since the I *Sce* I restriction analysis had been clearly suggestive of an *in trans* targeted line and the other conflicting results were not conclusive in their challenge of this assignment.

To determine the outcome of SSR events the HAT^R clones that were generated from the CAT-B3 and CAT-C10 lines were profiled for their drug resistance phenotypes and then Southern blot analysis performed on their DNA.

5.2 Experimental Outline

Cre-mediated SSR was conducted in cell lines previously established as targeted either *in cis* or *in trans*, as described in Chapter 3. The *in cis* line chosen for further study was CAT-B3 and the *in trans* line CAT-C10. For each cell line 2×10^8 cells were electroporated with either 50 μ g or 100 μ g of BAC DNA and in each case with 50 μ g of the Cre expression plasmid, CP#99.

Following electroporation cells were plated and cultured for 24 hours, viable cells were then counted and plated at densities of 5×10^7 or 1×10^7 and HAT selection added to the culture media. In addition, 2 plates with 1×10^7 cells had blasticidin instead of HAT added to the culture media. The blasticidin selection marker in the BAC is *floxed*, it was predicted that this marker would be deleted with efficiency should the BAC encounter Cre recombinase activity. Therefore, the plates with blasticidin selection alone represent the frequency with which cells have incorporated the BAC in the absence of Cre recombinase. Additionally, the resultant blasticidin^R clones were analysed by Southern blot to determine whether the BAC targets the alpha-

globin locus by homologous recombination (in the absence of Cre) or whether it integrates at random into the ES cell genome by illegitimate recombination.

After 5 days in selection colonies appeared and 20 days after electroporation colonies were picked into a 96 well plate. The clone numbers for each experiment are shown in Fig. 5.3.

Cells were cultured for a further 3 days and then split 1:3 into fresh 96 well plates. One of the three plates was frozen immediately, another was cultured to perform the drug resistance profiling and the cells of the remaining plate used for DNA preparation. The clones in the plate designated for drug resistance profiling were cultured to an adequate density before being split by replica plating into six 96 well plates. The culture media in each replica plate was supplemented with various drug selections and resistance was scored after a time interval appropriate to the selection in question, as recorded in Fig. 5.4.

Following the determination of the drug resistance profile for each clone, the genomic arrangement around the *hprt* minigene was investigated by Southern blot analysis. All of the viable clones underwent a primary screen for reconstruction with the UpRP, which detects fragments that are indicative of *hprt* minigene reconstruction, as detailed in Fig. 5.5. The results of the Southern blot analysis for clones generated from the CAT-B3 line are shown in Fig. 5.6 and for the CAT-C10 in Fig. 5.7. A number of clones were then selected for further characterization by Southern blot analysis.

The arrangement of the region external to the exchanged interval, that is upstream of the *loxP* site, was determined by Southern blot for a selected number of clones and the results of these analyses are shown in Fig. 5.8.

The predicted deletion of the hygromycin selection marker to leave a single *lox511* site and absence of the *HSVtk* gene downstream of the *lox511* site and the globin structural genes, was determined by flanking probes internal and external to the

Fig. 5.3 A table of the number of HAT resistant clones from each transfection

HAT ^R clones from CAT-B3 (<i>in cis</i> line)			HAT ^R clones from CAT-C10 (<i>in trans</i> line)		
	Colony numbers			Colony numbers	
Plate Density	50 µg BAC DNA	100 µg BAC DNA	Plate Density	50 µg BAC DNA	100 µg BAC DNA
5 x 10 ⁷	2	4	1 x 10 ⁷	4	4
	4	7		3	1
	3	6		2	4
	4	7		2	4
1 x 10 ⁷	0	2		4	0
	2	3		1	1
	0	3		3	0
	1	2		1	7
Total	16	31	Total	20	21
Frequency of HAT ^R clones	0.7 x10 ⁻⁷	1.3 x10 ⁻⁷	Frequency of HAT ^R clones	2.5 x10 ⁻⁷	2.6 x10 ⁻⁷

Fig 5.4a Resistance phenotypes of clones derived from the CAT-B3 line

	G418	Hyg.	Bsd.	Ganc.	HAT		G418	Hyg.	Bsd.	Ganc.	HAT
A1	S	S	S	R	R	C6	S	S	S	R	R
A2	S	S	S	R	R	C7	S	S	S	R	R
A3	S	S	S	R	R	C8	S	S	S	R	R
A4	S	S	S	R	R	C9	S	S	S	R	R
B1	S	S	S	R	R	D1	S	S	S	R	R
B2	S	S	S	R	R	D2	S	S	S	R	R
B3	S	S	S	R	R	D3	S	S	S	R	R
B4	S	S	S	R	R	D4	S	S	S	R	R
B5	S	S	S	R	R	D5	S	S	S	R	R
B6	S	S	S	R	R	D6	S	S	S	R	R
B7	S	S	S	R	R	D7	S	S	S	R	R
B8	S	S	S	R	R	D8	S	S	S	R	R
B9	S	S	S	R	R	D9	S	S	S	R	R
B10	S	S	S	R	R	D10	S	S	S	R	R
B11	S	S	S	R	R	D11	S	S	S	R	R
B12	S	S	S	R	R	D12	S	S	S	R	R
C1	S	S	S	R	R	E1	S	S	S	R	R
C3	S	S	S	R	R	E2	S	S	S	R	R
C4	S	S	R ?	R	R	E3	S	S	S	R	R
						*F1	R	R	R	R	S

Fig 5.4b Resistance phenotypes of clones derived from the CAT-C10 line

	G418	Hyg.	Bsd.	Ganc.	HAT		G418	Hyg.	Bsd.	Ganc.	HAT
A1	S	S	no data	R	R	B11	S	S	S	S	R
A2	S	S	no data	R	R	B12	S	S	S	R	R
A3	S	S	no data	S	R	C1	S	S	S	R	R ?
A4	S	S	no data	S	R	C2	S	S	S	R	R ?
A5	S	S	no data	R	R	C3	S	S	S	R	R
A6	S	S	no data	R	R	C4	S	S	S	R	R
A7	S	S	no data	R	R	C5	S	S	S	R	R
A8	S	S	no data	R	R	C6	S	S	S	R ?	S
A9	S	S	no data	R	R	C7	S	S	S	R ?	R ?
A10	S	S	no data	R	R	C8	S	S	S	R	R
A11	S	S	no data	R	R	C9	S	S	S	R	R
A12	S	S	no data	R	R	C10	S	S	S	R	R
B1	S	S	S	R	R	C11	S	S	S	R	R
B2	S	S	S	R	R	C12	S	S	S	R	R
B3	S	S	S	R	R	D1	S	S	S	R	R
B4	S	S	S	R	R	D2	S	S	S	R ?	R
B5	S	S	S	R	R	D3	S	S	S	S	S
B6	S	S	S	R	R	D4	S	S	S	R	S
B7	S	S	S	R	R						
B8	S	S	S	R	R	*D10	S	S	R	R	S
B9	S	S	S	R	R	*D11	R	R	R	R	S
B10	S	S	S	R	R	*D12	R ?	S	R	R ?	S

Fig. 5.4 records the drug resistance phenotypes for clones derived from the electroporation of pBAC14567/TK/3'HPRT into either the CAT-B3 or CAT-C10 double-targeted ES cell lines. Abbreviations: G418 = G418; Hyg. = hygromycin; Bsd. = blasticidin; Ganc. = ganciclovir; HAT = HAT. All of the clones were selected primarily in HAT supplemented media, with the exception of those marked with (*), which were selected in blasticidin supplemented media (clone F1 from the CAT-B3 line and clones D10, D11 and D12 from the CAT-C10 line). S denotes that the clone was sensitive to the selection, whilst R indicates that the clone is resistant. Some of the clones gave ambiguous results and are annotated with R ?, which signifies that most of the cells were sensitive, but that a small number continued to grow in the drug supplemented media.

exchange, as outlined in Fig. 5.9. The results of the analysis on selected CAT-B3 and CAT-C10 derived clones of the DnRP are shown in Fig. 5.10.

The reconstruction of the *hprt* minigene was further investigated by screening a selected number of clones with a 3'*hprt* probe internal to the reconstructed minigene (gift from Melville Richardson), the results are shown in Fig. 5.11a. The same selected clones were then screened using a neomycin probe (gift from Stephen Meek), which was the same probe used to detect the I *Sce* I fragment on the PFGE blot to assign *in cis* and *in trans* targeting (see section 3.2.1). The results of this analysis are also shown in Fig. 5.11b. For both of these analyses the position of the probes and restriction sites are illustrated in Fig. 5.5 (as above). Finally, Fig. 5.11b shows the results of the DnLP hybridisation of clones generated from the CAT-B3 and CAT-C10 lines, whose DNA was *Kpn* I restricted (see Fig. 5.9 for predicted fragment sizes).

5.3 Analysis of the efficiency with which HAT^R clones were generated from the CAT-B3 and CAT-C10 lines

It was predicted that the efficiency of generating HAT^R clones from the CAT-B3 would be significantly higher than for the CAT-C10 line (see section 5.1). However, the data presented in Fig. 5.3 suggest that in fact, HAT^R clones are generated with greater frequency from the CAT-C10 line. Although these figures represent just two experiments (one with 50 µg of BAC DNA and one with 100 µg) in each case the frequency of generating HAT^R clones from the CAT-C10 line was higher. Clearly, this is contrary to prediction and to the reports of comparisons of RMCE with single *lox* site strategies, with intervals of less than 10 kb.

It was noted however, that the CAT-B3 line shows an increase in the generation of HAT^R colonies when 100 µg of BAC DNA is used compared with 50 µg. In contrast, the CAT -C10 line does not demonstrate the same increase in the frequency of HAT^R clones. This may suggest that the SSR events in the CAT-C10 line have reached an optimum level for generating HAT^R clones, whilst the frequency of generating

HAT^R clones from the CAT-B3 line could be further increased, possibly by the introduction of more BAC DNA or by increasing the efficiency with which the BAC is delivered to the ES cells. This is consistent with the observations of integration/excision systems that reach an optimum level of integrated product due to the competing reverse excision reaction, compared with the dynamics of an RMCE system for which the reverse reaction is thought to be negligible (Bethke and Sauer, 1997) (refer back to sections 3.4.1 and 5.1). However, in order to verify this tentative hypothesis the ES cells would have to be electroporated with quantities of BAC DNA in excess of 100 µg. The introduction of such quantities of DNA by electroporation may be detrimental to cell survival and therefore make this line of investigation impractical.

The blasticidin resistant clones were generated from each line with an average frequency of 1×10^{-7} and 1.5×10^{-7} , for the CAT-B3 and CAT-C10 lines respectively.

5.4 Analysis on the drug resistance phenotypes of clones generated from the CAT-B3 and CAT-C10 lines

The predicted outcome of SSR between pBAC14567/TK/3'HPRT and the *in cis* targeted CAT-B3 line was that RMCE would occur and for the *in trans* line, which was CAT-C10, the generation of a co-integrand would persist (see Fig. 5.2). Before Southern blotting was performed on the clones generated from the CAT-B3 and CAT-C10 lines it was anticipated that their genetic arrangement could be predicted from their drug resistance profiles. Fig. 5.4 records the drug resistance phenotypes from the HAT^R and blasticidin^R clones generated from each line.

As predicted in Fig. 5.2, following RMCE the predicted drug resistance phenotype for the CAT-B3 generated clones was: HAT^R; ganciclovir^R; blasticidin^S; G418^S and hygromycin^S. As Fig. 5.4 records, almost 100% of the HAT^R clones from the CAT-B3 line gave the predicted drug resistance phenotype (with the exception of

clone C4, for which sensitivity to blasticidin was not absolute). This initial observation suggested that the large interval RMCE had been achieved.

Assuming that a co-integrant forms (see Fig. 5.2), the predicted drug resistance phenotype for the CAT-C10 generated clones was: HAT^R; ganciclovir^S; blasticidin^S; G418^R and hygromycin^S. As Fig. 5.4 shows, the status of blasticidin resistance for some of the clones was not established (no data). However, with respect to the other clones and markers, all of the clones demonstrate the predicted sensitivity to hygromycin and blasticidin.

The generation of co-integrant structures from the CAT-C10 line predicts that the *HSVtk* gene would be introduced to the genome following SSR via the *loxP* or *lox511* sites to integrate the entire BAC (see Fig. 5.2). Furthermore, the neomycin gene would be retained in co-integrant structures. It was therefore unexpected that the clones generated from the CAT-C10 line were all sensitive to G418 selection and furthermore that the majority were also resistant to ganciclovir selection (except clones C6, C7 and D2, which although not obviously ganciclovir resistant, did show some cell death and may therefore represent co-integrant clones). The resistance phenotypes for the CAT-C10 clones were obviously contradictory to the predicted results, however it was anticipated that Southern blot analysis would clarify the genetic arrangement with respect to this line.

The blasticidin resistant clones generated from each line were all resistant to blasticidin and sensitive to HAT, which is consistent with the prediction that these clones represent the random integration of the BAC. Clone F1, which was generated from the CAT-B3 line was resistant to all the other selections, which would indicate that Cre has not been active in this clone, since the floxed hygromycin and blasticidin genes would most likely be deleted if Cre had been present. Clone F1 can be most readily explained by a random integration event that involved parts of the BAC, but not including the *HSVtk* gene. Clone D11 from the CAT-C10 line has an equivalent phenotype to F1 and can therefore be assumed to represent an equivalent event. Clones

Fig 5.5 Predicted restriction fragments and probe locations around *loxP* for *in cis* and *in trans* HAT^R RMCE generated clones

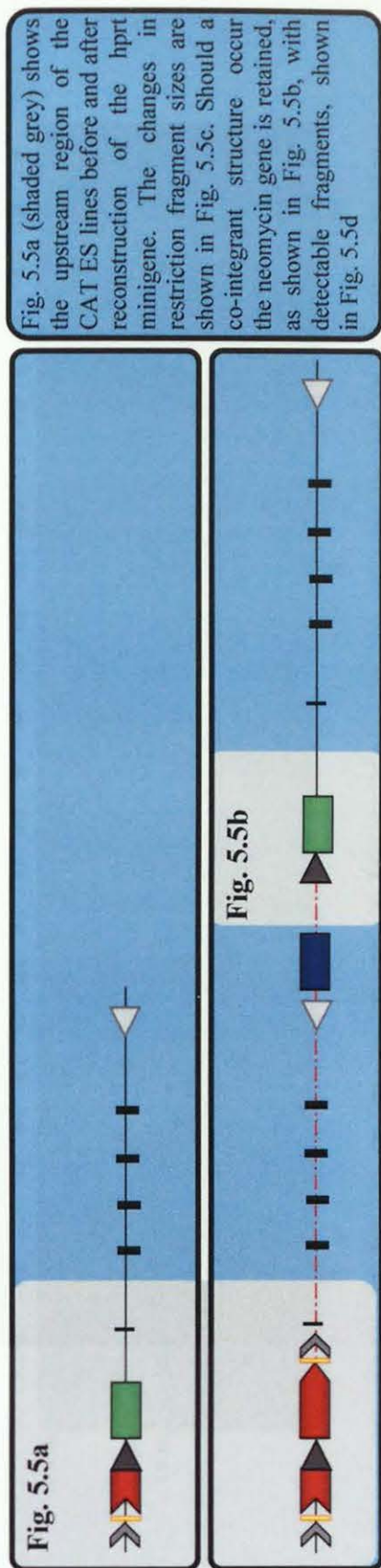
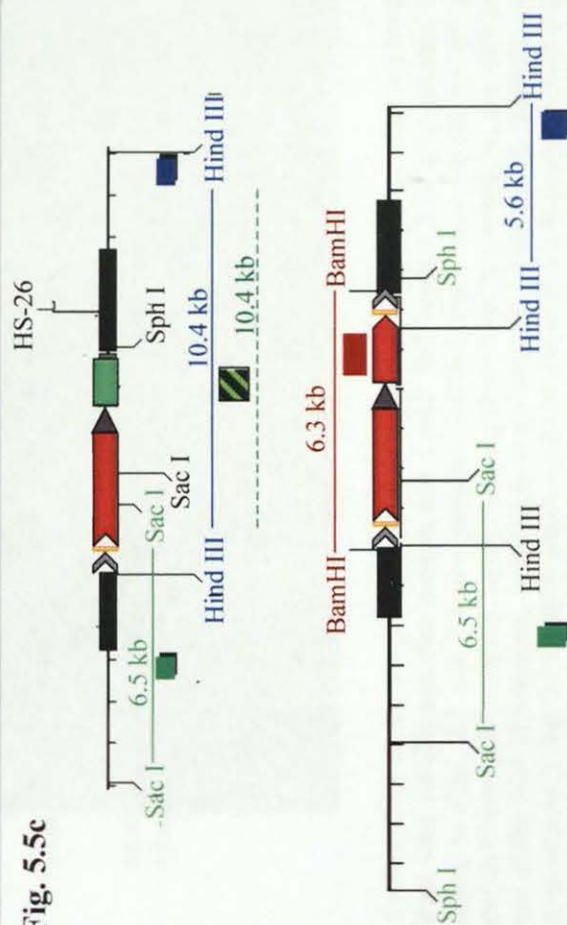


Fig. 5.5c



Key

- 1 kb
- Upstream Right Probe (UpRP)
- Upstream Left Probe (UpLP)
- Restriction fragment recognised by 3'hprt probe
- Restriction fragment recognised by 3'hprt probe
- Restriction fragment recognised by UpRP
- Restriction fragment recognised by UpLP

Fig. 5.5d

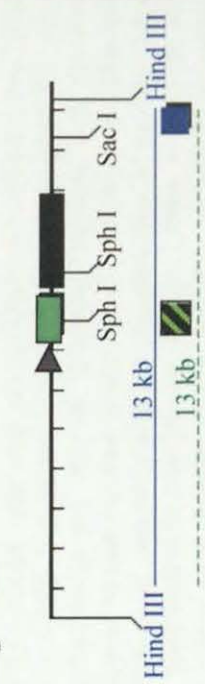
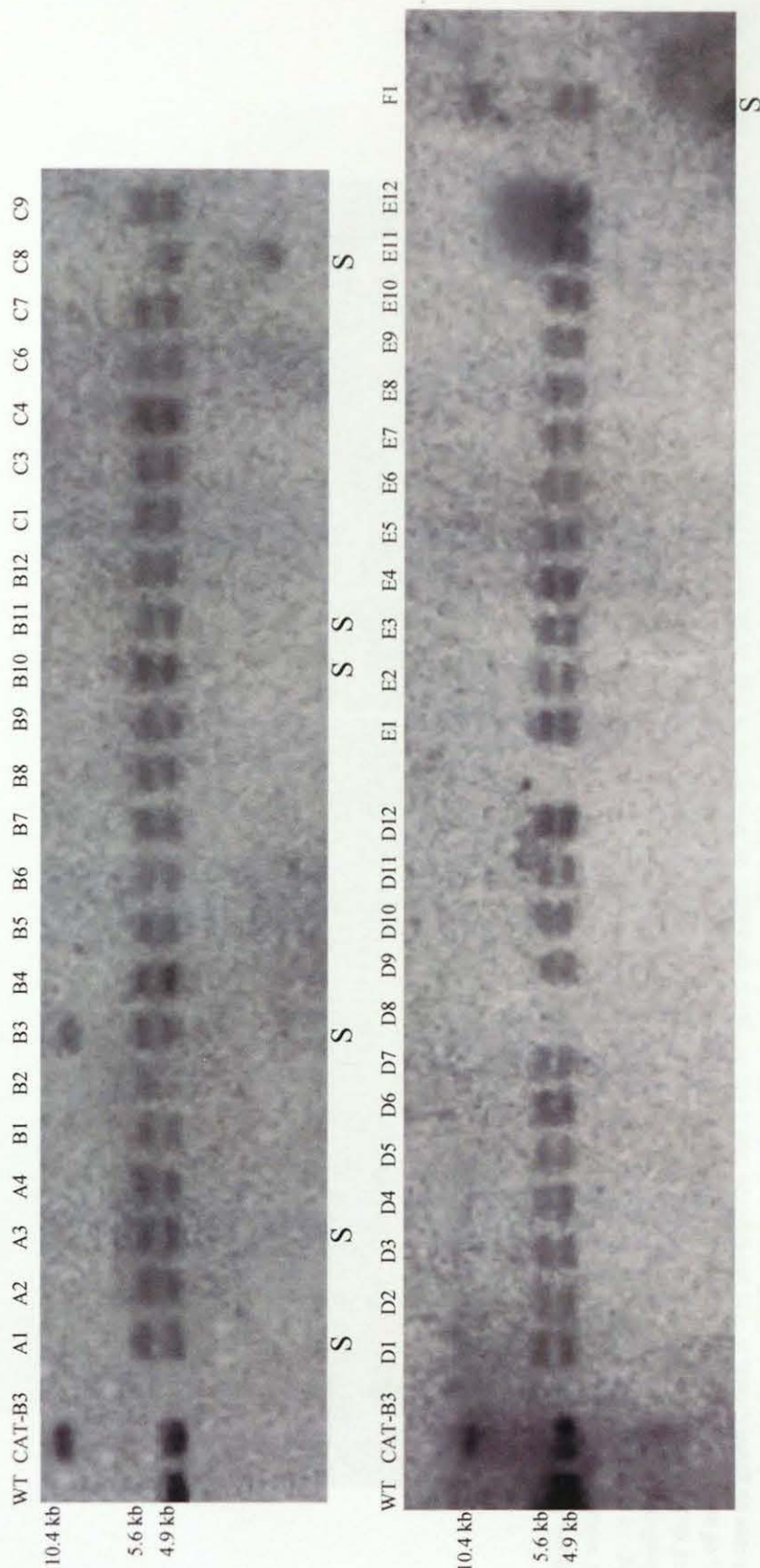
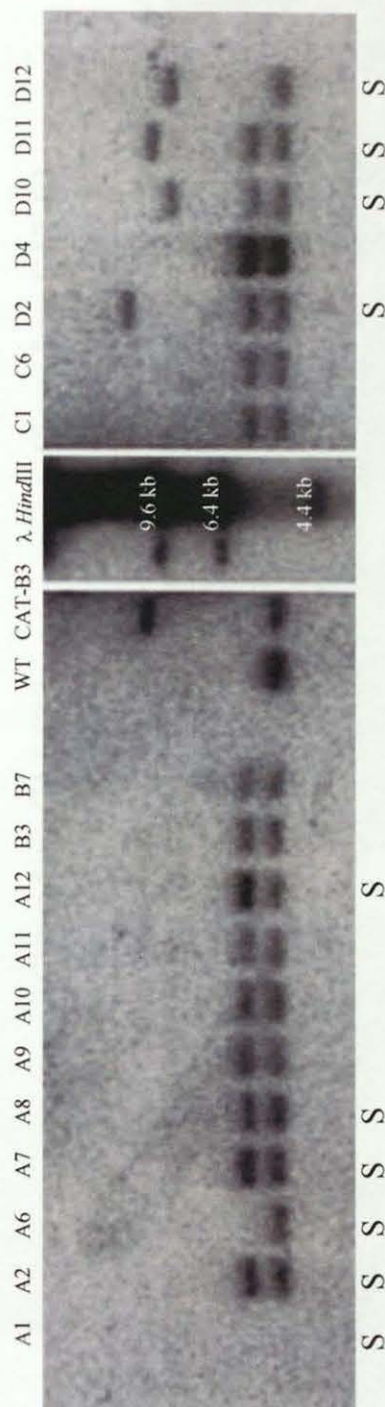


Fig. 5.6 Southern blot analysis on clones generated from the CAT-B3 line, with UpRP



The Southern blot shows the primary screen of the clones resulting from electroporation of CAT-B3 ES cells with pBAC14567/TK/3'HPRT. The DNA was restricted with *HindIII* and then probed with UpRP. WT denotes E14Tg2a genomic DNA. CAT-B3 is included to represent double-targeted ES cell lines. According to Figs. 2.7 and 5.5, the expected banding sizes were 4.9 kb for an untargeted locus and 10.4 kb following targeting with the *frt/I* *Sce* I/5'hprr/*loxP*/neomycin construct. Upon reconstruction of the hprr minigene, the 10.4 kb fragment should decrease in size to a 5.6 kb fragment. Therefore the absence of the 10.4 kb fragment, in addition to the presence of the 5.6 kb fragment was taken to be consistent with the reconstruction of the hprr minigene. However, of note is that the 5.6 kb fragment would be detected as a fragment internal to BAC integrated at random into the genome. Clones A to E were selected in HAT supplemented media, whilst clone F1 was selected for with blasticidin. 45 of 48 clones show banding that is indicative of HPRT reconstruction, including the F1 clone, which was selected in blasticidin. The clones marked with an 'S' were selected for further analysis by Southern blotting to determine their genetic arrangement and were: A1; A3; B3; B10; B11; C8; F1.

Fig. 5.7 Southern blot analysis on clones generated from the CAT-C10 line, with UpRP



The Southern blot shows the primary analysis of the clones resulting from electroporation of CAT-C10 ES cells with pBAC14567/TK/3'HPRT. The DNA was restricted with *HindIII* and then probed with UpRP. **WT** denotes E14Tg2a genomic DNA. **CAT-B3** is included to represent double-targeted ES cell lines. According to Figs. 2.7 and 5.5 the expected banding sizes were 4.9 kb for an untargeted locus and 10.4 kb following targeting with the *frt/I* *Sce* I/5'hprt/*loxP*/neomycin construct. Upon reconstruction of the hprt minigene, the 10.4 kb fragment should decrease in size to a 5.6 kb fragment. Therefore the absence of the 10.4 kb fragment, in addition to the presence of the 5.6 kb fragment was taken to be consistent with the reconstruction of the hprt minigene. However, of note is that the 5.6 kb fragment would be detected as a fragment internal to BAC integrated at random into the genome. In addition, since CAT-C10 is the *in trans* targeted clone, it was anticipated that a co-integrant structure, as described in Fig. 5.5b, would be detected from the presence of a 13 kb band. All the clones were selected for in HAT supplemented media, except clones D10, D11 and D12, which were selected for in blasticidin. The clones marked with an 'S' were selected for further analysis by Southern blotting to determine their genetic arrangement and were: A1; A2; A6; A7; A8; A12; D2; D10; D11; D12.

Fig. 5.8 Checking the integrity of the locus upstream of the RMCE interval for selected clones generated from the CAT-B3 and CAT-C10 lines, with UpLP

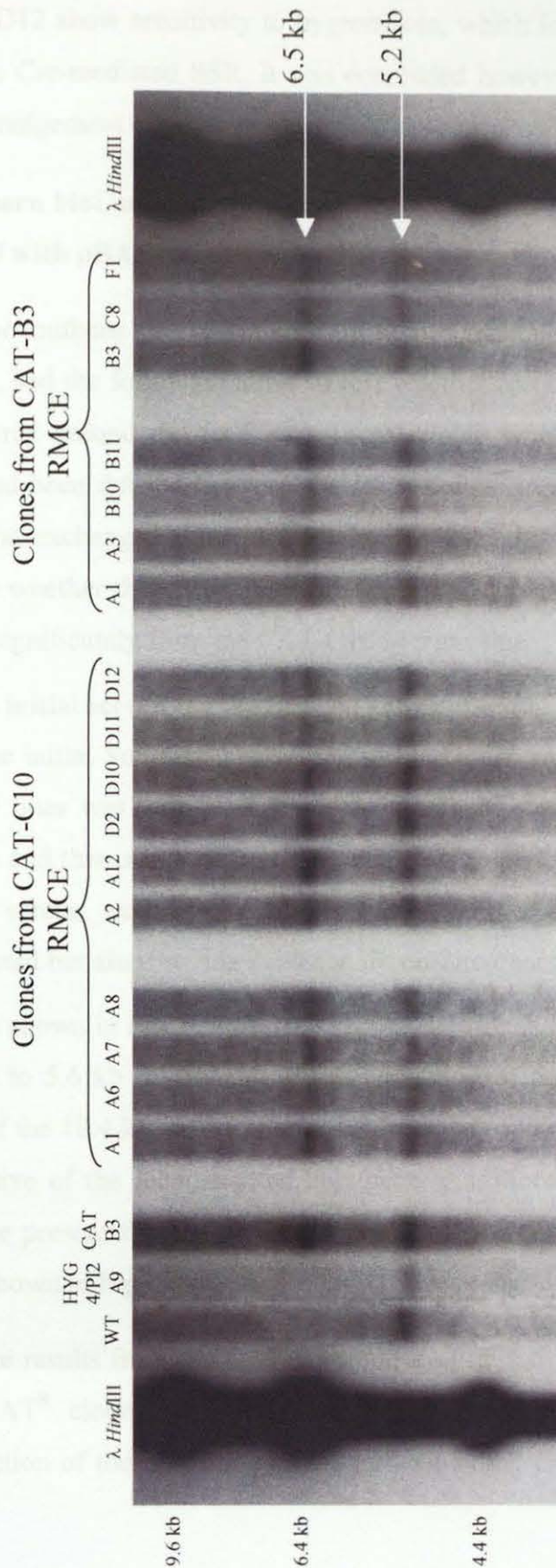


Fig. 5.8 shows the investigation of selected clones generated from the *in cis* (CAT-B3) and *in trans* (CAT-C10) ES cell lines following their electroporation with CP#99 and pBAC14567/TK/3'HPRT. WT denotes E14Tg2a genomic DNA. HYG-4/PL2 is the cell line targeted with the downstream ES cell targeting construct, which was used to generate the CAT lines. CAT-B3 is included to represent double-targeted ES cell lines. The CAT-B3 generated clones A1, A3, B10, B11, B3 and C8 were all HAT^R and gave results consistent with the reconstruction of the hprt minigene (see Fig. 5.6). The CAT-B3 generated F1 clone was blastocidin^R and is proposed to represent the random integration of pBAC14567/TK/3'HPRT in the absence of Cre. The CAT-C10 derived clones D10, D11 and D12 were all HAT^R, however clone D2 generated a fragment that suggested that this clone represented a co-integrant clone (see Fig. 5.7). The CAT-C10 derived clones D10, D11 and D12 were blastocidin^R and are proposed to represent the random integration of pBAC14567/TK/3'HPRT in the absence of Cre. The DNA from the selected clones was restricted with *SacI* and then probed with UpLP. The *SacI* sites that generate the fragment to which UpLP hybridises are outwith the RMCE interval and so the Southern blot data for all of the clones should remain unchanged from the original double targeted clones, of 5.2 kb from the untargeted chromosome and 6.5 kb from the targeted chromosome. As shown, all of the clones screened demonstrate fragments equivalent to the CAT-B3 double-targeted clone.

D10 and D12 show sensitivity to hygromycin, which indicates that these clones have undergone Cre-mediated SSR. It was concluded however that an accurate idea of the genetic arrangement of these clones would only be possible by Southern blotting.

5.5 Southern blot analysis of clones derived from electroporation of CAT-B3 and CAT-C10 with pBAC14567/TK/3'HPRT

The southern blot analysis, on the clones generated from the CAT-B3 and CAT-C10 lines, had the following aims: to test whether reconstruction of the *hprt* minigene had occurred around the *loxP* site; to determine whether the hygromycin selection marker had been deleted between the *lox511* sites; to determine whether the regions outwith the exchanged interval were intact, both upstream and downstream; and to determine whether there was any evidence for co-integrant structures from either line, but most significantly from the CAT-C10 *in trans* line.

5.5.1 The initial screen for the desired recombinants

The initial Southern blot screen of the clones generated from the CAT-B3 and CAT-C10 lines was with UpRP, which is a probe that hybridises with alpha-globin sequences and that is internal to the exchange interval (see Fig. 5.5). It was anticipated that this screen would not only indicate whether the *hprt* minigene had been reconstructed but also provide evidence for co-integrant structures.

As shown in Fig. 5.5 the 10.4 kb *Hind*III fragment from the upstream target site decreased to 5.6 kb following the reconstruction of the *hprt* minigene, therefore the absence of the 10.4 kb fragment and the presence of the 5.6 kb fragment were taken to be indicative of the reconstructed *hprt* gene. Additionally, if a co-integrant structure were to be present the UpRP should detect a 13kb fragment spanning the neomycin gene, as shown in Fig. 5.5d.

The results from the CAT-B3 line (see Fig. 5.6) demonstrate that 44 out of 47 of the HAT^R clones gave the predicted fragment sizes that are consistent with reconstruction of the *hprt* minigene. The blasticidin^R clone CAT-B3/F1 also indicates

the presence of an 5.6 kb fragment, which is thought to arise from the randomly integrated BAC sequences and the 10.4 kb band from the upstream target site that has not undergone SSR. These results for the F1 clone are further consistent with its assignment as a random integrant clone.

The results from the CAT-C10 line (see Fig. 5.7) demonstrate that 13 out of 15 HAT^R clones tested generate Southern blot data that are consistent with the reconstruction of the *hprt* minigene. However, just one clone, CAT-C10/D2, generated a fragment that would suggest that this clone represents a co-integrant. This result for CAT-C10/D2 is consistent with its drug resistance phenotype (see Fig. 5.4). If this clone does indeed represent a co-integrant the *HSVtk* gene should be present and thereby cause ganciclovir sensitivity and although the majority of the cells for the D2 line were sensitive, a small number were found to be persistently viable.

The CAT-C10/D11 blasticidin^R clone, which gave a similar drug resistance phenotype to the CAT-B3/F1 clone, also generates equivalent Southern blot data to the CAT-B3/F1 clone, which further confirms that these two clones represent equivalent random integrant events.

The other two blasticidin^R clones, CAT-C10/D10 and CAT-C10/D12 gave unusual Southern blot data: both clones generate a fragment of about 9 kb, which can not be explained and whereas clone CAT-C10/D10 demonstrates the presence of the 5.6 kb fragment that arises from randomly integrated BAC sequences, clone CAT-C10/D12 does not. These results can not be explained at this time.

It was anticipated that not all the HAT^R clones generated from the CAT-B3 and CAT-C10 line would generate the predicted results from Southern blot analysis, due to the usual occurrence of anomalous results that are experienced with any genetic manipulation experiment. However, the fact that the majority of the CAT-B3 generated clones did generate the anticipated fragment sizes was indicative that these clones probably harboured the predicted and desired genetic arrangement.

5.5.2 Checking the integrity of the genome outwith the RMCE interval

To verify that the region upstream of the reconstructed *hprt* minigene was intact a selected number of clones were restricted with *Sac* I and then probed with UpLP. The fragments that should result are illustrated in Fig. 5.5. The *Sac* I sites that generate the fragment to which UpLP hybridises are outwith the RMCE interval and so the Southern blot data for all of the clones should remain unchanged from the original double targeted clones. Therefore a fragment of 5.2 kb was expected from the untargeted chromosome and 6.5 kb from the targeted chromosome. The data in Fig. 5.8 demonstrate that all of the clones screened demonstrate these expected fragments and are therefore consistent with the region upstream of the exchanged interval being intact.

Similarly, in order to verify that the region downstream of the interval outwith the *lox511* site was intact, the same clones were investigated using a southern blot analysis described in Fig. 5.9, with the DnRP. According to Fig. 5.9, the *floxed* hygromycin gene is deleted during the desired large interval RMCE event. There is however no concomitant change in fragment size since the *Bam*HI sites that border the *lox511* site persist. Therefore the predicted *Bam*HI fragments sizes are ~10 kb for the untargeted locus, ~7.3 kb for the targeted locus and following deletion of the hygromycin, the persistence of the same ~7.3 kb fragment (in addition to the fragment that arises from the untargeted chromosome). Fig. 5.10 demonstrates that the region downstream of the *lox511* site has the predicted genetic arrangement for all of the HAT^R clones tested.

Unusually, this probe does not detect fragments arising from the random integration of the BAC for the blasticidin^R clones. However, since all of the blasticidin^R clones were ganciclovir^R (see Fig. 5.4), this result is consistent with their drug resistance phenotype, which suggests that the BAC has randomly integrated but not included the HSV θ k gene.

Fig 5.9 Predicted restriction fragments recognised by DnLP and DnRP for *in cis* and *in trans* HAT^R RMCE generated clones

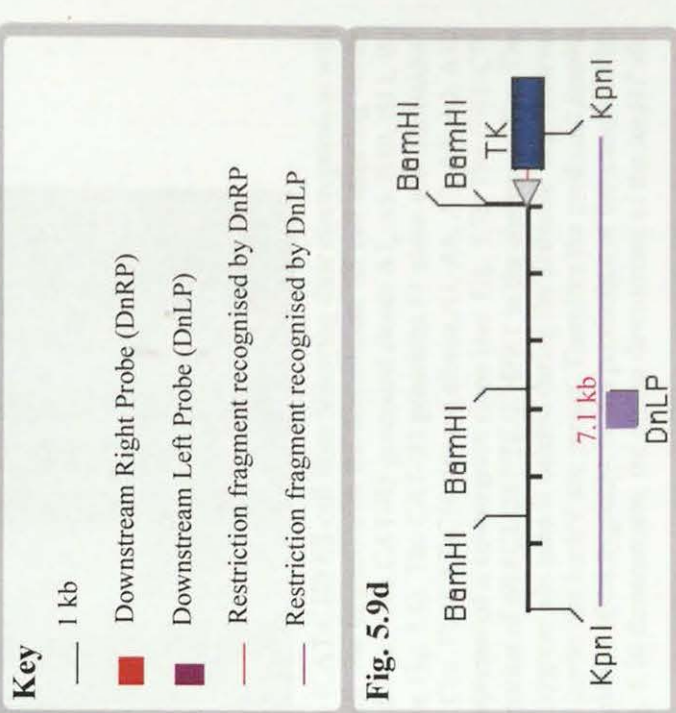
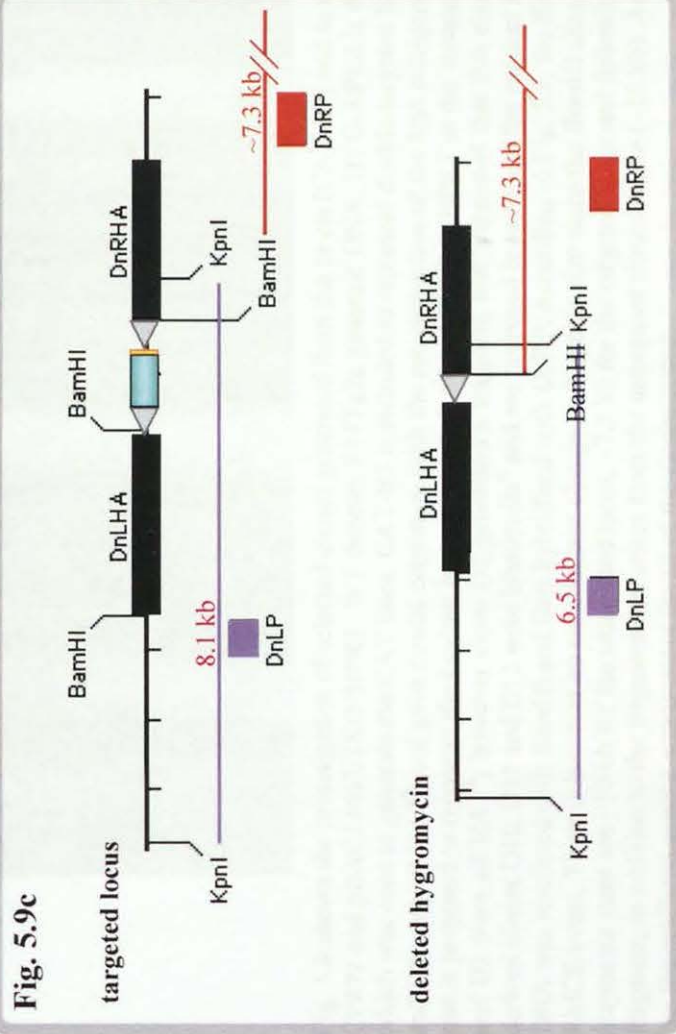
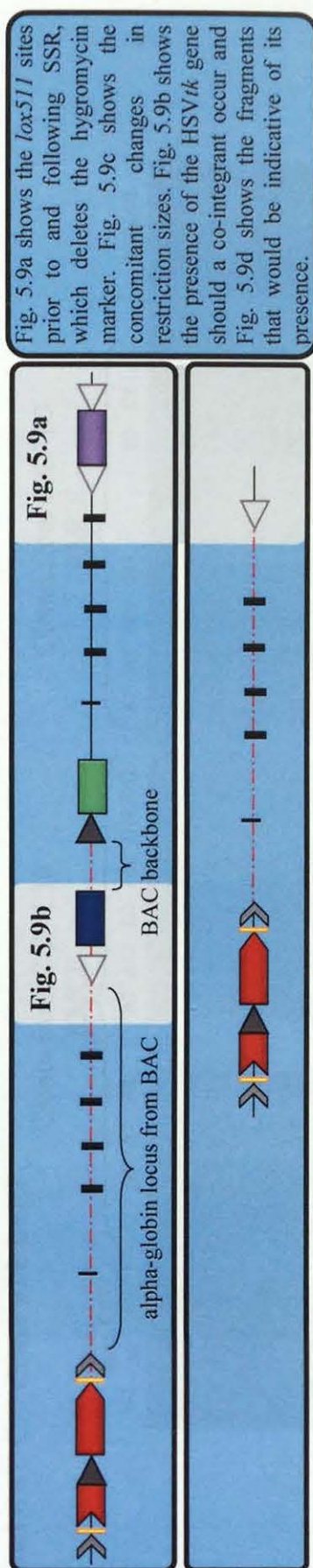


Fig. 5.10 Checking the integrity of the locus downstream of the RMCE interval for selected clones generated from the CAT-B3 and CAT-C10 clones with DnRP

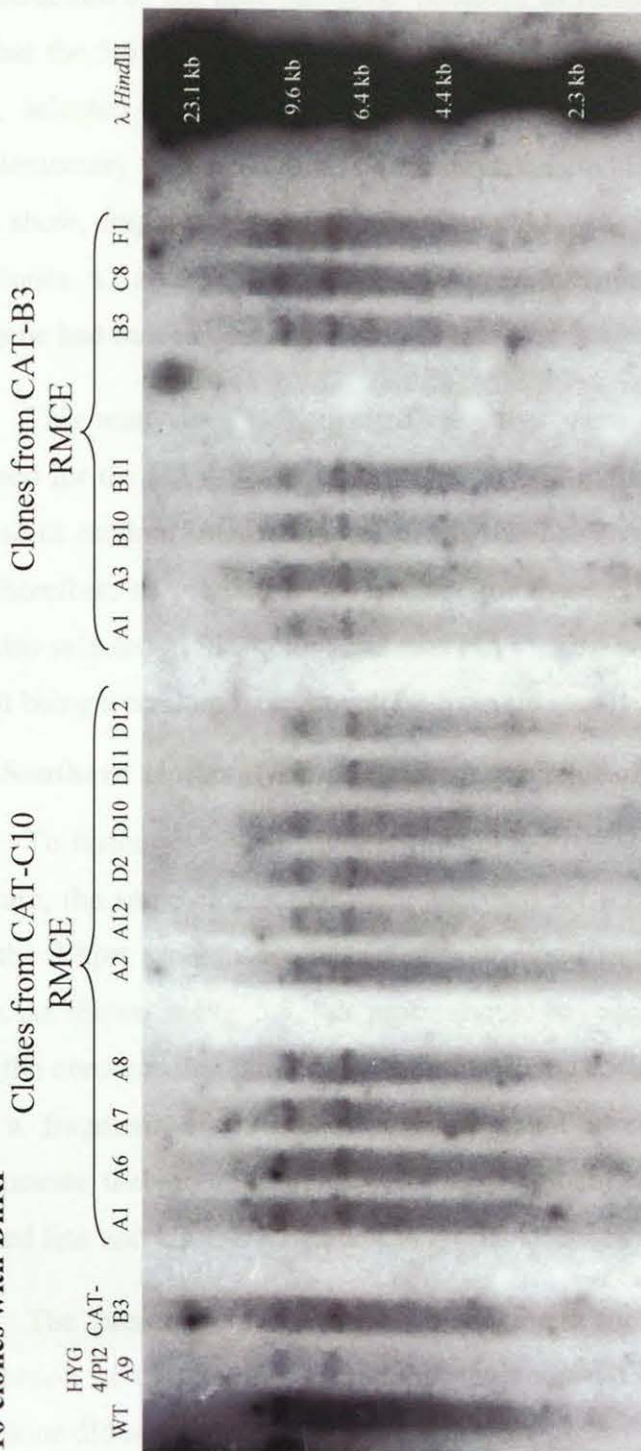


Fig. 5.8 shows the investigation of selected clones generated from the *in cis* (CAT-B3) and *in trans* (CAT-C10) ES cell lines following their electroporation with CP#99 and pBAC14567/TK/3'HPRT. WT denotes E14Tg2a genomic DNA. HYG-4/PL2 is the cell line targeted with the downstream ES cell targeting vector which was used to generate the CAT lines. CAT-B3 is included to represent double-targeted ES cell lines. The CAT-B3 generated clones A1, A3, B10, B11, B3 and C8 were all HAT^R and gave results consistent with the reconstruction of the hprt minigene (see Fig. 5.6). The CAT-B3 generated F1 clone was blasticidin^R and is proposed to represent the random integration of pBAC14567/TK/3'HPRT in the absence of Cre. The CAT-C10 derived clones A1, A6, A7, A8, A2, A12 and D2 were all HAT^R, however clone D2 generated a fragment that suggested that this clone represented a co-integrant clone (see Fig. 5.7). The CAT-C10 derived clones D10, D11 and D12 were blasticidin^R and are proposed to represent the random integration of pBAC14567/TK/3'HPRT in the absence of Cre. The CAT-C10 DNA was restricted with *Bam*HI and then hybridised with DnRP. According to Fig. 5.9, the *lox* sites that border the *lox511* site persist. Therefore the predicted large interval RMCE event. There is however no concomitant change in fragment size since the *Bam*HI sites that border the *lox511* site persist. Therefore the predicted large interval fragments sizes are ~10 kb for the untargeted locus, ~7.3 kb for the targeted locus and following deletion of the hygromycin, the persistence of the same ~7.3 kb fragment, in addition to the fragment that arises from the untargeted chromosome (~10 kb). As Fig. 5.10 demonstrates, the region downstream of the *lox511* site has the predicted genetic arrangement for all of the clones tested.

5.5.3 Reconstruction of the *hprt* minigene

Thus far, analysis with the UpRP around the *loxP* site was consistent with the reconstruction of the *hprt* minigene, however to further confirm that this was the case and that the 5.6 kb band was not generated by sequences from a randomly integrated BAC, selected clones were digested with *Bam*HI and hybridised with a probe complementary to the 3' region of the reconstructed *hprt* minigene. As the data in Fig. 5.11a show, fragments generated by clones A1, A3, B10 and B11 from CAT-B3 line and clones A7 and A8 from CAT-C10 line confirmed that reconstruction of the *hprt* minigene had indeed occurred.

This analysis also generated data that were consistent with previous results obtained for the CAT-B3/F1 and CAT-C10/D11 clones, which are believed to harbour equivalent random integrations of the BAC. The reconstruction of the *hprt* minigene was therefore, as predicted, not evident for these clones. The CAT-C10/D10, which was also selected in blasticidin and was HAT^S, also generated data that were consistent with it being a random integrant clone from this analysis.

5.5.4 Southern blot analysis to establish evidence of co-integrant structures

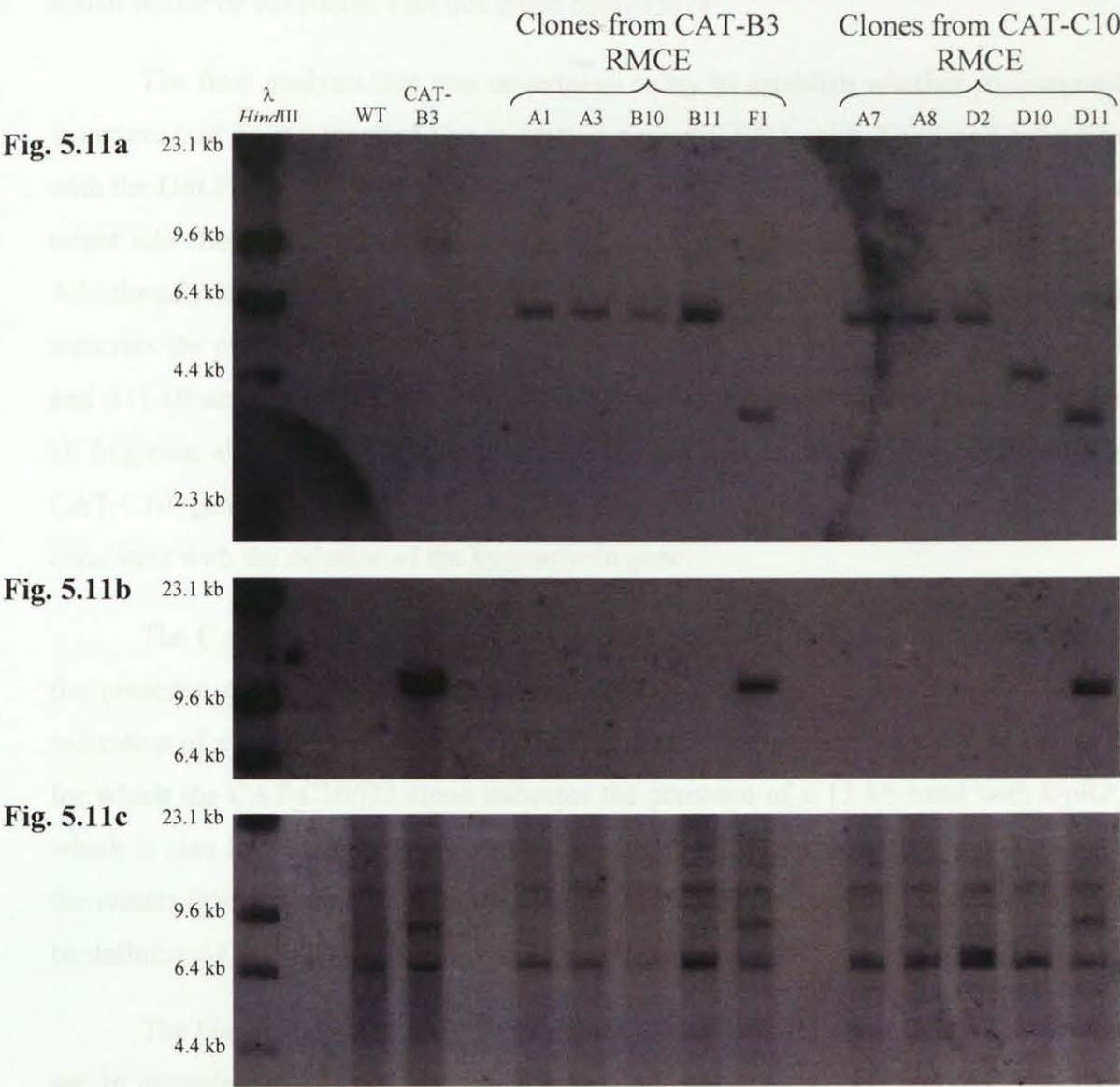
To further establish whether the CAT-C10/D2 clone did harbour a co-integrant structure, the same clones that were analysed for reconstruction of the *hprt* minigene with the 3'*hprt* probe, were restricted with *Hind*III and hybridised with the neomycin probe. As shown in Fig. 5.5, this probe should have hybridised with a 10.4 kb fragment from the chromosome targeted with the *frr*/I *Sce* I/5'*hprt/loxP*/neomycin construct and with a fragment of 13 kb if a co-integrant structure was present. Clones that demonstrate the presence of the 10.4 kb fragment are: the control CAT-B3 double-targeted line and CAT-B3/F1 and CAT-C10/D11.

The presence of the 10.4 kb fragment for clones F1 and D11 is further verification of these clones as random integrants, it was surprising therefore that the D10 clone did not generate the same fragment.

Fig. 5.11 Southern blot analysis on selected clones generated from the CAT-B3 and CAT-C10 lines, with 3'hppt probe, neomycin probe and DnLP

Fig. 5.11 shows the Southern blot analysis on selected clones generated from both the CAT-B3 and CAT-C10 lines. WT denotes E14Tg2a genomic DNA and the other control DNA was from the CAT-B3 double-targeted ES cell line. The CAT-B3 generated clones A1, A3, B10, and B11 were HAT^R and gave results consistent with the reconstruction of the hppt minigene (see Figs. 5.6 and 5.8). The CAT-B3 generated F1 clone was blasticidin^R and was proposed to represent the random integration of pBAC14567/TK/3'HPRT in the absence of Cre. The CAT-C10 derived clones A7, A8 and D2 were all HAT^R, however clone D2 generated a fragment that suggested that this clone represented a co-integrant clone (see Fig. 5.7). The CAT-C10 derived clones D10 and D11 were blasticidin^R and are proposed to represent the random integration of pBAC14567/TK/3'HPRT in the absence of Cre. Fig. 5.11a shows the hybridisation of *Bam*HI restricted genomic DNA hybridised with the 3'hppt probe (gift from Melville Richardson). A 6.3 kb fragment indicates reconstruction of the hppt minigene, which is evident for clones A1, A3, B10 and B11 generated from CAT-B3 and clones A7, A8, and D2 for the CAT-C10 generated clones. Fig. 5.11b shows the hybridisation of *Hind*III restricted genomic DNA hybridised with the neomycin probe (gift from Stephen Meek). This probe should hybridise with a 10.4 kb fragment from the chromosome targeted with *frt/I Sce I/5'hppt/loxP*/neomycin and with a fragment of 13 kb if a co-integrant structure is present. Clones that demonstrate the presence of the 10.4 kb fragment are: the control CAT-B3 double-targeted line and clones F1 and D11. No clones demonstrate the presence of the 13 kb fragment. Fig. 5.11c shows the hybridisation of *Kpn*I restricted genomic DNA hybridised with the DnLP. Fig. 5.9 predicts that the targeted fragment of 8.1 kb should be reduced to 6.5 kb following deletion of the hygromycin marker, as evident for all of the CAT-B3 HAT^R clones. The CAT-B3/F1 clone indicates the presence of an 8.1 kb fragment, which is consistent with its assignment as a random integrant clone. The CAT-C10 clones A7 and A8 show fragments consistent with the deletion of the hygromycin gene. The CAT-C10 clone D2 shows the presence of an additional fragment, which may be the 7.1 kb fragment that arises from the *Kpn*I fragment caused by the presence of the HSVtk gene in a co-integrant structure. This result is in accordance with that in Fig. 5.7, for which the CAT-C10/D2 clone indicates the presence of a 13 kb band with UpRP, which is also indicative of a co-integrant structure. The CAT-C10 blasticidin resistant clones should give identical fragment sizes to the CAT-B3 double-targeted line, however only clone D11 gives this result, clone D10 appears to have deleted the hygromycin gene. In addition the DnLP probe is detecting a background fragment, that appears to be about 15 kb and is present in all the lanes, including the controls.

Fig. 5.11 Southern blot analysis on selected clones generated from the CAT-B3 and CAT-C10 lines, with 3'hppt probe, neomycin probe and DnLP



None of the clones demonstrated the presence of the 13 kb fragment. Most surprising was the result for the CAT-C10/D2 clone, which was indicated as harbouring a co-integrant from data in Fig. 5.7. This result suggests that either CAT-C10/D2 is not a co-integrant or that the neomycin gene has somehow been deleted, which would be consistent with this clone being G418^S.

The final analysis that was undertaken to try to establish whether co-integrant structures had been generated was to restrict genomic DNA with *Kpn* I and hybridise with the DnLP. Fig. 5.9 predicts that a fragment of 8.1 kb arising from the downstream target site should be reduced to 6.5 kb following deletion of the hygromycin gene. Additionally, this probe recognises a 7.1 kb fragment should a co-integrant occur and indicates the presence of the HSV ι k gene. The CAT-B3 generated clones A1, A3, B10 and B11 all demonstrate the absence of the 8.1 kb fragment and the presence of the 6.5 kb fragment alone, which is consistent with the deletion of the hygromycin gene. The CAT-C10 generated clones A7 and A8, also show equivalent fragments that are consistent with the deletion of the hygromycin gene.

The CAT-C10/D2 clone that was initially predicted to be a co-integrant shows the presence of an additional fragment, which may be the 7.1 kb fragment that is indicative of a co-integrant structure. This result is in accordance with that in Fig. 5.7, for which the CAT-C10/D2 clone indicates the presence of a 13 kb band with UpRP, which is also indicative of a co-integrant structure. However this result conflicts with the results from the analysis with the neomycin probe, therefore CAT-C10/D2 can not be definitively established to represent a co-integrant structure.

The blasticidin^R clones CAT-B3/F1 and CAT-C10/D11 again generate data that are in accordance with the data in Fig. 5.8 and their resistance profiles. The data generated here for these two clones would further suggest that the HSV ι k gene appears to be absent from the randomly integrated BAC sequence, since a fragment of 7.1 kb is not detected in these clones.

Clone CAT-B3/D10 appears to have deleted the hygromycin gene, which is consistent with its resistance phenotype.

The DnLP probe appears to detect a fragment of about 15 kb, however since this is present in each line, including the controls, it is presumed to be background and therefore not informative of the genetic arrangement of the clones.

5.6 Summary

The analyses on the HAT^R clones generated from the CAT-B3 in *cis* targeted line gave results that were consistent with the achievement of large interval RMCE in these clones.

The generation of blasticidin resistant clones are indicative that the BAC DNA will undergo random integration into the ES cell genome and this event is thought to occur if Cre recombinase activity is low or absent from these lines, as determined by the persistence of floxed selection genes in these clones.

The confidence of achieving the large interval RMCE would be strengthened had the CAT-C10 clone generated data that were consistent with the predicted generation of co-integrant structures from this line. However, all of the HAT^R clones generated from the CAT-C10 line appear to have deleted the neomycin gene and furthermore most do not indicate the presence of the HSVtk gene (except CAT-C10/D2), which is contradictory to the predicted co-integrant structure that should result for this line, but consistent with the CAT-C10 line being targeted *in cis*.

The analysis of the CAT-C10 clone, as described in Chapter 3, was not conclusive in establishing this clone as targeted *in trans* and this could explain the anomalous results obtained from this line in this chapter. Alternatively, it is possible that the CAT-C10 line is a genuine *in trans* line but events other than SSR, such as homologous recombination are occurring to give rise to the anomalous results for this line.

If SSR and homologous recombination are occurring within the CAT-C10 line, it is plausible that these events would also occur in the CAT-B3 line. Therefore, Chapter 6 discusses the possibility for a combination of homologous recombination

and SSR events to occur during large interval RMCE experiments and the effects that these would have on achieving large interval RMCE.

Chapter 6

DISCUSSION

6.1 Introduction

The aim of the experiments described within this thesis, were to test whether a strategy that employs the rationale of RMCE could be applied to the study of whole genes, gene clusters and large gene loci. To this end, the application of RMCE with larger intervals than previously reported, required demonstration.

The alpha-globin locus was chosen as an example of a locus spanning a genomic region of tens of kilobases, since the exchange of an interval of this size should be representative of technology that may be applied to encompass the average gene locus (Lander et al., 2001; Venter et al., 2001). In addition, the exchanged interval includes structural and regulatory sequences, whose co-ordinated interaction provides a highly informative model to further study mechanisms of gene-regulation.

The Southern blot analysis, presented in the Chapter 5, sought to determine whether the genetic arrangement at each end of the exchanged interval was consistent with the arrangement of a locus that had undergone the predicted SSR events. For the *in cis* line tested (CAT-B3) this was anticipated to be a Cre-mediated large interval RMCE event. The data gathered from the HAT^R clones generated from the CAT-B3 (*in cis*) line are consistent with the achievement of the large interval RMCE event in these clones, on three points.

First, the reconstruction of the *hprt* minigene was definitively established, which was absolutely dependent on the SSR between the *loxP* site in the BAC and the *loxP* site targeted to the ES cell chromosome upstream of the locus.

Second, following the reconstruction of the *hprt* minigene, it was predicted that the entire BAC molecule would be integrated to the ES cell chromosome. However, the absence of both the *HSVtk* gene and the hygromycin gene from the CAT-B3 generated HAT^R clones would suggest that SSR around the *lox511* has occurred to resolve this structure and cause the large interval RMCE event.

Thirdly, if large interval RMCE has occurred the HAT^R clones should not harbour the neomycin gene and this was also established by resistance profiling and Southern blotting.

Alternative interpretations of the data are possible and would be consistent with data gathered thus far reflecting alternative recombination events, other than SSR to generate RMCE proposed thus far. These events are discussed later in section 6.3.

However, this chapter first describes the analysis that would be performed to determine that the entire interval that has been exchanged intact, this analysis is particularly relevant to the application of large interval RMCE, for example to introduce mutations within the exchanged interval.

6.2 Characterisation of the exchanged interval

The exchange described within this thesis was mediated between an ES cell line and a BAC harbouring an isogenic cloned DNA insert. It was therefore impossible by restriction site polymorphisms between the two molecules to determine whether the interval in the BAC had exchanged into the interval of the ES cell intact.

Another strategy to characterise the exchanged interval would take advantage of the fact that the BAC and ES cell interval were identical by determining that the interval following SSR events was unchanged. This could be achieved by designing ten

to twenty overlapping long range PCR amplifications of 5 to 10 kb, throughout the interval. The results of this analysis on the HAT^R clones generated from the CAT-B3 line and controls of untargeted ES cell and BAC DNA would be predicted to demonstrate that each generates the same size PCR products. This analysis would therefore be able to determine any re-arrangements made to the interval of 500 bp or more but would not be useful to definitively determine re-arrangements smaller than this.

In order to establish whether the strategy developed within this thesis is useful to facilitate the mutational analysis of large genomic regions (as intended) the definitive experiment would be to introduce mutations to the BAC interval and then determine whether all of the intended mutations had been delivered by PCR or Southern blotting techniques. In addition, this experiment would establish whether the intact interval had been exchanged.

Alternatively, a BAC that contained non-isogenic DNA could be used to mediate the exchange and H. Wallace is currently undertaking this experiment. The experiment involves exchanging the human alpha-globin sequence from a BAC into the alpha-globin ES cell locus. This experiment thereby provides human specific restriction fragment length polymorphisms and human specific PCR primer sites that can be used to characterise the interval that is anticipated to undergo exchange.

6.3 The effects of undesirable recombination on large interval RMCE

The introduction of the BAC to ES cells could produce a variety of genomic re-arrangements: alone or combined, homologous recombination (HR), illegitimate recombination (IR) and SSR of one, or more, BACs introduced to the ES cell could generate these genomic re-arrangements.

The BAC vector used to deliver the exchanged interval has, due to its cloned insert, a large region of homology with the alpha-globin loci of the ES cell genome. The possibility that this homology could mediate HR between the BAC and ES cell

genome therefore requires consideration. Moreover, the propensity for HR, or IR, to occur in combination with SSR needs exploring, since the outcome of these events may be highly detrimental to the efficiency and application of this strategy.

In order to isolate RMCE events a selection strategy that reconstructs the *hprt* minigene and excludes the *HSVtk* gene was therefore implemented. In addition, following clonal selection, Southern blot analysis further elucidated the genetic arrangement of HAT resistant (HAT^R) clones.

The isolation of HAT^R clones is absolutely dependent on the reconstruction of the *hprt* minigene, which in turn implicates a SSR event around the *loxP* sites. As proposed in the previous chapter, the reaction may proceed via the production of a co-integrant structure, which may have then resolved by a second SSR event. However the possibility exists that the HAT^R clones may be generated by alternate recombination events. Briefly, HAT^R clones may theoretically be generated by a combination IR followed by SSR (see Fig. 6.1) or by a combination of HR either preceding or following SSR (see Fig. 6.2). The subsequent drug resistance profiling and Southern blot analysis of HAT^R clones was therefore performed in an order to identify whether the HAT^R phenotype represents the desired RMCE event.

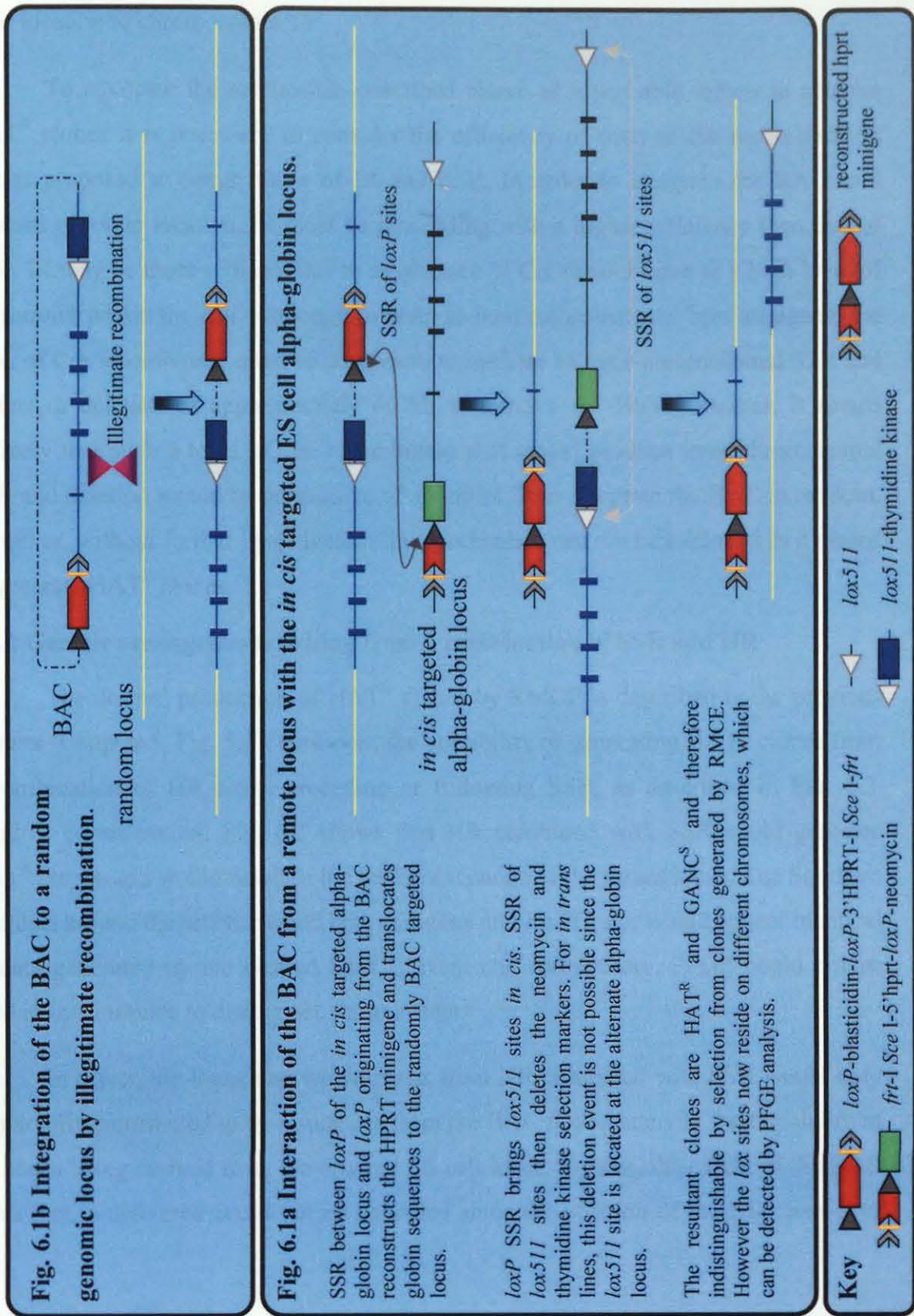
6.3.1 Genetic arrangements arising from IR followed by SSR

The production of HAT^R clones from *in cis* targeted lines, by the integration of the BAC to a random genomic location by IR followed by SSR is shown in Fig. 6.1. As illustrated, reconstruction of the *hprt* minigene and deletion of the *HSVtk* selection marker occur, events that are also outcomes of the desired RMCE event. Therefore, by Southern blot analysis alone it is not possible to distinguish RMCE from IR followed by SSR. However, following RMCE the *lox* sites are predicted to remain *in cis*, whereas for IR followed by SSR the *lox* sites become positioned *in trans*. In order therefore to exclude clones with a HAT^R phenotype generated by IR followed by SSR, two analyses could be performed: the *in cis* or *in trans* arrangement of the *lox* sites could be determined by performing PFGE, followed by Southern blot analysis; or FISH

Fig. 6.1 Proposed model to generate HAT^R clones from *in cis* targeted ES cell lines by illegitimate recombination followed by SSR

The diagram illustrates the production of HAT^R clones by integration of the BAC to a random genomic location by IR, followed by resolution by SSR to produce a genomic arrangement that would appear to be a RMCE by the southern blot analysis performed.

Fig. 6.1 Proposed model to generate HAT^R clones from *in cis* targeted ES cell lines by IR followed by SSR



analysis could determine whether alpha-globin sequences were properly restricted in their location to chromosomes 11.

To advocate the mechanism described above as a probable means to achieve HAT^R clones it is necessary to consider the efficiency of each of the recombination events proposed to occur, those of IR and SSR. In order to integrate the BAC to a random genomic location, IR must be proceeding with a higher efficiency than that of SSR. IR may be more efficient due to an absence of Cre recombinase or a high level of IR activity within the cell. Although in order to then reconstruct the *hprt* minigene, the level of Cre recombinase must be competent to mediate an inter-chromosomal SSR and further, a deletion of approximately 60 kb to remove the *HSVtk* marker. It seems unlikely that such a level of Cre recombinase that would produce inter-chromosomal SSR and deletion would be permissive of an initial IR to integrate the BAC at random. However, without further investigation this mechanism can not be excluded as a means to generate HAT^R clones.

6.3.2 Genetic arrangements arising from a combination of SSR and HR

The desired production of HAT^R clones by RMCE is described in the previous chapter (Chapter 5, Fig. 5.1). However, the possibility of generating HAT^R clones from a combination of HR either preceding or following SSR, as described in Fig. 6.2, requires consideration. Fig. 6.2 shows that HR combined with SSR could generate HAT^R clones and would result in the partial exchange of the *floxed* locus. The Southern blot data around the reconstructed *hprt* minigene and *lox511* site would appear identical to data generated by the desired RMCE event and furthermore, PFGE could not be used in this instance to distinguish these events.

In effect, the locus that would result from HR combined with SSR would only be partially contributed to by sequences from the BAC *floxed* interval, the remainder of the locus being derived from the original ES cell locus. Further, the extent of the BAC locus that is delivered could not be specified since the position of the HR cross-over

can not be controlled. Clearly this hinders the application of the strategy since extensive screening would be required to determine the extent of the exchanged region. The following section discusses the likelihood of HR occurring. However the definitive experiment that would establish whether this is the case has already been described in section 6.2.

6.3.3 Conditions under which the BAC could participate in HR

In order to assess whether the BAC may participate in HR it is necessary to first discuss conditions under which HR has been demonstrated to occur. As discussed in Chapter 1, HR integrates gene-targeting vectors to their target sequence by virtue of the sequence homology to the target within the vector.

An obvious difference between the BAC DNA used here and gene-targeting vectors is that the latter is linearised as a course of most targeting protocols. Linearisation introduces double-strand break termini to the DNA molecule and these are thought to promote HR.

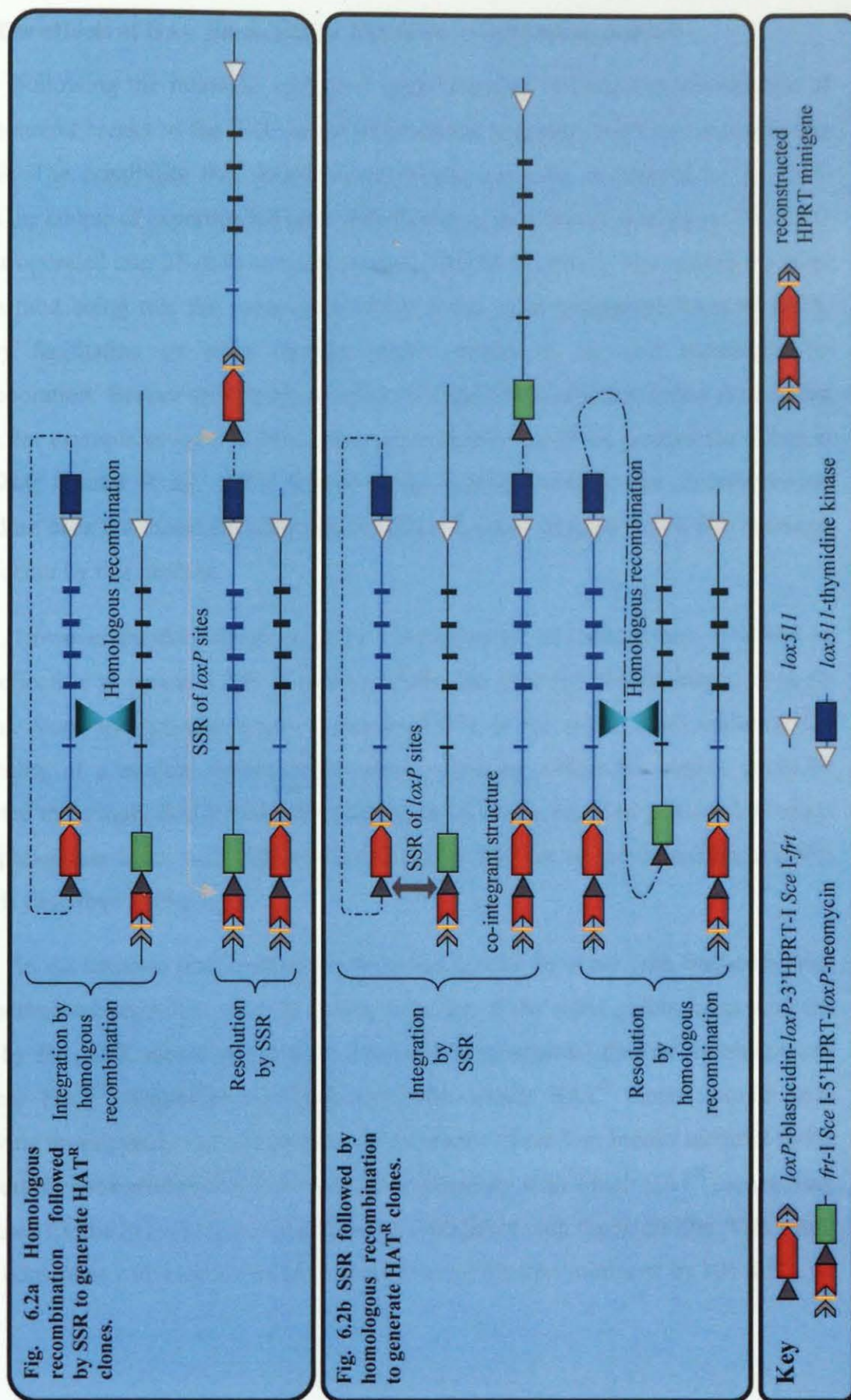
It would seem logical therefore to assume that since the BAC is electroporated into the ES cells as a super-coiled circular molecule, without breaks, that the possibility for HR to occur would be greatly reduced. The integration of the circular BAC by HR to the alpha-globin locus could be predicted therefore to be unlikely. Further, the efficiency of SSR may diminish the occurrence of a low level of HR by being a more efficient event.

It remains however to establish whether HR with circular molecules, of BAC size and homology, could occur with frequency. In addition, the incidence of double-strand breaks to the BAC during the experimental procedure may further increase its ability to participate in HR and these points are addressed below.

Fig. 6.2 A diagram to illustrate the generation of HAT^R clones by the combined events of SSR and homologous recombination with *in cis* targeted ES cell lines

The diagram illustrates the production of HAT^R clones by a combination of HR and SSR between the BAC and ES cell DNA, which have been engineered to generate a RMCE event. As illustrated, the combination of HR and SSR in any order results in the partial exchange of the RMCE interval. Although the HR and SSR events are shown for an *in cis* line, the equivalent events could occur in an *intrans* line with the chromosome targeted with the *frt*/I *Sce* I/ 5'hpvt/*loxP*/neomycin sequences.

Fig. 6.2. A diagram to illustrate the generation of HAT^R clones by the combined events of SSR and homologous recombination with in cis targeted ES cell lines



6.3.4 The effects of BAC linearisation due to experimental procedure

Following the rationale applied to gene-targeting vectors, the introduction of double-strand breaks to the BAC could be predicted to greatly increase its propensity for HR. The possibility that double-strand breaks could be introduced to the BAC within the course of experimental procedure therefore requires consideration. The BAC is electroporated into ES cells as a supercoiled, circular molecule. The rationale behind this method being that the super-coiled form is the most compacted form of DNA, thereby facilitating its entry through pores created in the cell membrane by electroporation. Breaks are known to occur in large DNA molecules due to shearing forces, for example as experienced during microinjection or DNA preparation (Chen et al., 1994; Gnirke et al., 1993; Schedl et al., 1993). However the electroporation procedure does not create shearing forces and so it seems unlikely that DNA breakage could occur by this method.

However should a break occur by electroporation or other means, it would be most effective to promote HR if present within the genomic cloned insert, or at its termini. Since the genomic insert makes up 85 % of the entire BAC molecule the probability of a random double-strand break occurring within this region could be predicted to be high. BACs with such double-strand breaks could be predicted to target the alpha-globin locus with high efficiency due to the size of the homology and this event is described in Fig. 6.2a.

In the scenario that random double-strand breaks do occur with frequency, due to experimental procedure, then following targeting of the alpha-globin locus with the BAC by HR, SSR should occur with efficiency to reconstruct the hprt minigene and generate HAT^R. Therefore the frequency with which HAT^R clones would arise following these events would represent the incidence of random breaks incurred to the BAC during the experimental procedure. The efficiency with which HAT^R clones were generated for the *in trans* line was higher than compared with the *in cis* line. This result is not consistent with clones that have reconstructed the hprt minigene by HR followed

by SSR, since this event should generate equal efficiencies for both the *in cis* and *in trans* line. Therefore it is predicted that HR due to the incidence of double-strand breaks incurred to the BAC is not a frequent event.

6.3.5 The consequences of a combination of SSR followed by HR

In consideration of the points of the above discussion, the most likely alternate mechanism for the production of HAT^R clones is predicted to be SSR followed by HR, as depicted in Fig. 6.2b.

Assuming that the BAC molecule remains circular and thus HR occurs at a low frequency, SSR could then be predicted to be the dominant primary recombination event. To generate HAT^R clones SSR must occur via the *loxP* sites to reconstruct the *hprt* minigene and in doing so generate the co-integrand structure, as outlined in Fig. 6.2b. The issue pertinent to RMCE then becomes whether the co-integrand is resolved desirably, by a second SSR event between the *lox511* sites, or by HR. The Southern blot data obtained thus far are unable to distinguish these events, however the comparison of data generated from the HAT^R clones from the *in cis* and *in trans* lines may provide an indication of which event resolves the co-integrand.

38 HAT^R clones were generated from the *in cis* lines, none of which have been identified with Southern blotting to harbour a co-integrand structure. In contrast, 1 of the 14 clones analysed from the *in trans* generated HAT^R clones demonstrates the persistence of the co-integrand structure, the remaining clones appearing to have resolved, according to Southern blot analysis.

Assuming that all the *in trans* generated clones may resolve the co-integrand structure either by HR or SSR, the frequency of co-integrand clones from the total number provides an indication of the activity of HR. Therefore in the absence of a *lox511* site on the same chromosome HR appears to occur efficiently (13 of 14 clones) to cause the resolution of the co-integrand structure for the *in trans* clones. The possibility for a high incidence of HR to occur between *in cis* sequences is recorded at

other genomic locations (Baker, 1989; Ng and Baker, 1994), however the alpha-globin locus has not been recorded as a 'hot-spot' for homologous recombination.

A measure of the frequency of HR following SSR for the *in trans* generated clones would be obtained by determining the frequency with which CAT-C10/D2 clone was able to generate clones with a resolved genetic arrangement in the absence of Cre. This experiment is currently being undertaken by measuring the number of ganciclovir resistant clones that arise spontaneously from this clone. However this measurement may not reflect the level of HR that may be occurring for the clones generated from the *in cis* line since resolution by SSR is possible.

Assuming that either HR or SSR may resolve the co-integrant structure for the *in cis* generated clones (HR is active at an equivalent frequency for both the *in trans* and *in cis* generated clones) then it would be reasonable to predict that from a population of 38 HAT^R clones generated from the CAT-B3 line, at least 2 or 3 would exhibit a co-integrant genetic arrangement. Since the Southern blot data suggest that none of the *in cis* generated clones has a co-integrant structure, then either HR is active at a higher efficiency in these cells, which is unlikely, or resolution by other means is occurring, most likely by SSR via the *in cis* targeted *lox511* site. This scenario would suggest that in a number of HAT^R clones from this experiment the entire *floxed* interval has undergone exchange, as desired.

Furthermore, it may be that the effects of HR for the *in cis* generated clones are diminished due to the greater efficiency of the desired SSR resolution by *lox511*, compared with the *in trans* generated clones. The position of the *lox511* site on the same chromosome, for the *in cis* line, may promote SSR above HR as the dominant event which resolves the co-integrant structure. Should this be the case then the frequency of recovering clones that have undergone exchange of the entire *floxed* interval, within a population of HAT^R clones, would be increased. However, the distance between the *in cis* *lox* sites may affect whether SSR or HR is the dominant resolution even.

One of the aims of developing this strategy was to generate a method that could facilitate the functional dissection of a locus of interest. To this end, the ability to precisely and rapidly deliver any combination of mutations to a locus of interest was required. Currently, to direct the multiple mutagenesis of a locus with conventional gene-targeting methods requires multiple, sequential rounds of gene-targeting, a process which is both lengthy and possibly detrimental to the cell. In contrast, the strategy developed here seeks to facilitate the introduction of multiple mutations by a single RMCE event into a pre-disposed locus. The possibility for HR to occur would therefore severely hamper the application of this strategy in this respect, since not all the mutations engineered into the BAC *floxed* interval could be guaranteed to be exchanged into the *floxed* ES cell interval.

As outlined above, the HAT^R clones generated could represent a mixed population of clones harbouring various genomic arrangements at the locus of interest, with each having undergone varying degrees of exchange with the BAC. In order to implement the strategy described here to introduce mutations to a locus of interest, it would therefore be necessary to either evaluate the extent of the exchange, or implement further screening to identify the mutations transferred during the RMCE event.

In order to determine clones that have had all the required mutations delivered it would be possible to mark each mutation with a drug selectable marker. Therefore following the isolation of HAT^R clones it would be possible to further select clones, to identify those with the full complement of mutations on the basis of their drug resistance profile. However, this would compromise one of the main advantages of the strategy, which is the ability to position drug selectable markers and other undesirable sequences out with the region under investigation. Furthermore, it is possible to deliver point mutations alone to the BAC interval and then using this strategy it would be possible to deliver such mutations without the concomitant introduction of any other

sequences. At present the ability to manipulate the ES cell genome with such precision can not be achieved and is therefore a useful development.

A more appropriate method therefore would be to screen the DNA from all of the HAT^R clones for the mutations intended for delivery. The methods that could be used to do this were described earlier, in section 6.2. The importance of determining the extent of the exchange would be of specific relevance on applying this strategy to the RMCE of heterologous sequences. Although the propensity for HR to occur with heterologous sequences could be predicted to be lower than for the experimental conditions described here, there is evidence that non-isogenic DNA participates in HR (Deng C, 1992). In the case of heterologous sequence exchange, intensive screening for polymorphic sequences between species would therefore be required to determine the extent of the exchanged interval.

Finally, the generation of a population of HAT^R clones harbouring various genetic arrangements may be a fortuitous event, since with a single RMCE event it would be possible to generate a number of clones with varying mutational profiles. However, should a level of HR be undesirable, it may be possible to increase the efficiency of SSR within the system by providing more robust Cre expression within the cell. This could be achieved, for example, by optimising the ratio of Cre expression plasmid to the amount of electroporated BAC, or by using cell lines that have a highly inducible Cre recombinase gene, either from an episome or by the development of Cre expression cell lines.

Chapter 7

MODIFICATIONS TO THE STRATEGY AND APPLICATIONS

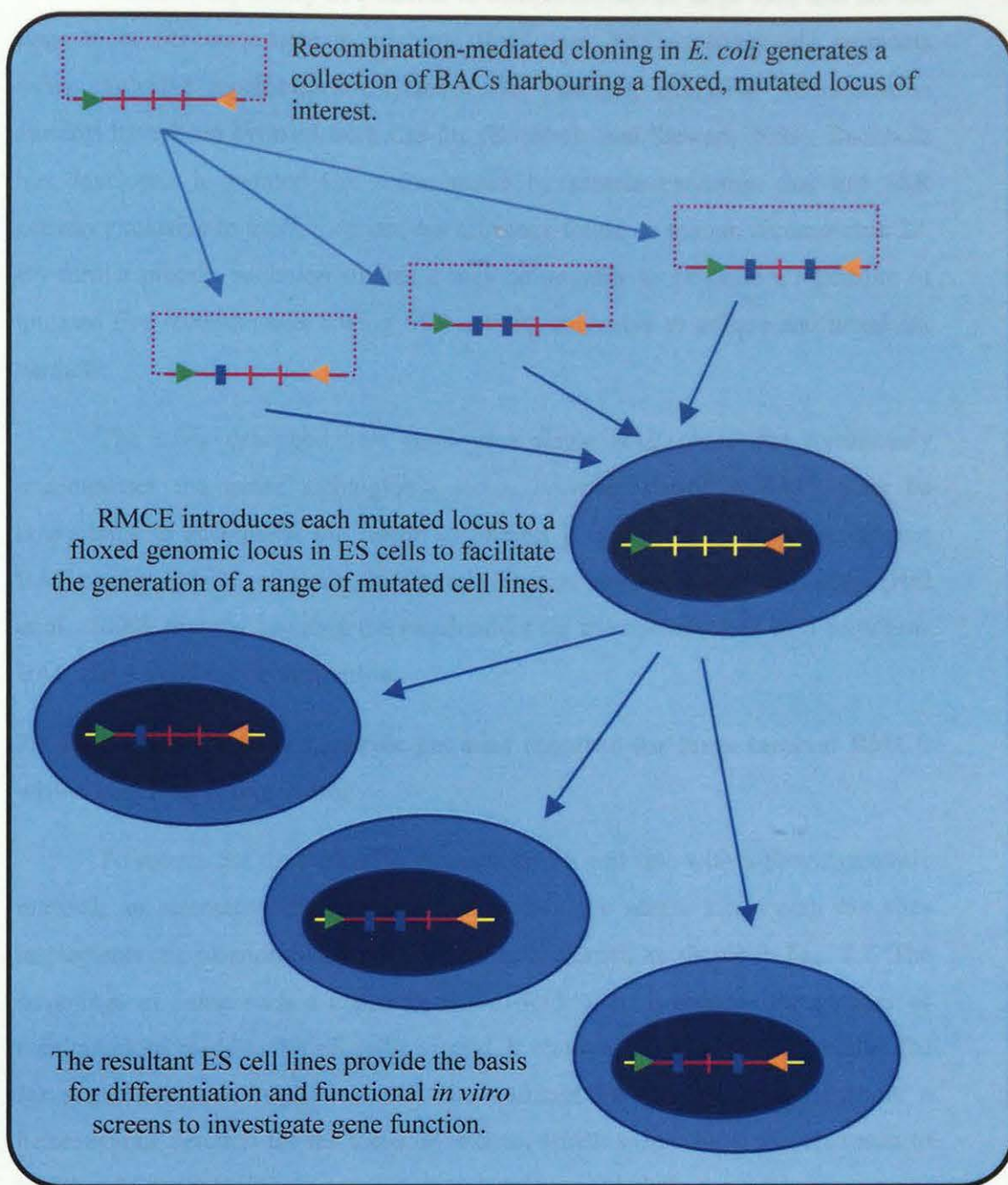
7.1 Modifications to the large interval RMCE strategy in order to expediate the experimental process

The strategy developed in this thesis was designed to facilitate the recursive introduction of mutations to a specified locus in order to perform functional genomics, see Fig. 7.1. By the addition of other technology to this strategy of large interval RMCE, it may be possible to flexibly extend this strategy to the study of any region of interest and expediate the study. The following section outlines some of the modifications that could be made in this respect.

7.2 Modifications to the large interval RMCE strategy to study gene loci that can not be accommodated by a single BAC molecule

The ability to exchange intervals of up to 300 kb, harboured in BACs, permits the exchange of most genes and gene loci, however, a requirement to exchange larger intervals with an equivalent strategy could be met using alternative vectors, for example HACs, MACs or whole chromosomes. At present, the ability to shuttle whole chromosomes by microcell-mediated transfer is being tested as a vector system for RMCE in order to facilitate the replacement of large regions of synteny between human and mouse in the murine ES cells using an equivalent strategy (Helen Wallace, pers. comms. Centre for Genome Research, Edinburgh University, 2001).

Fig. 7.1 A strategy to rapidly produce a variety of mutated ES cell lines



The diagram illustrates how a single ES cell line with a *floxed* locus of interest can be used to generate a variety of mutated cell lines. The mutations are introduced to the locus harboured in the BAC and then with a single RMCE event can be delivered to the ES cell *floxed* interval.

Alternatively it may be possible to compartmentalize large loci, that are too large to be accommodated by a single BAC, into BAC-exchangeable segments using exclusive recombination systems. For example, exclusive recombination systems have been evolved from Cre-*lox* (Buchholz and Stewart, 2001). Buchholz has developed a mutated Cre recombinase by protein evolution that has SSR activity exclusive to *loxH*, a variant *lox* sequence found on human chromosome 22. By further protein evolution studies it may be possible to generate a repertoire of mutated Cre recombinases having SSR activity exclusive to unique and novel *lox* variants.

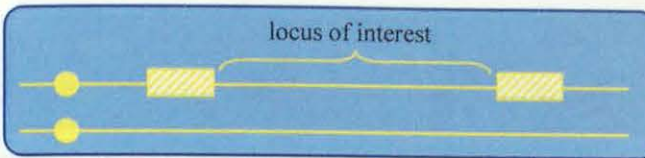
The study described here employs a single BAC clone that fortuitously encompasses the entire alpha-globin locus, however should a BAC clone be unavailable to encompass the region of interest it is possible to join overlapping BAC contigs (Mejia and Larin, 2000) and eliminate any unwanted insert DNA (Hill et al., 2000), thereby negating the requirement for the re-derivation of a fortuitous BAC clone by library construction.

7.3 Delivering the hetero-specific *lox* sites required for large interval RMCE with a single targeting vector

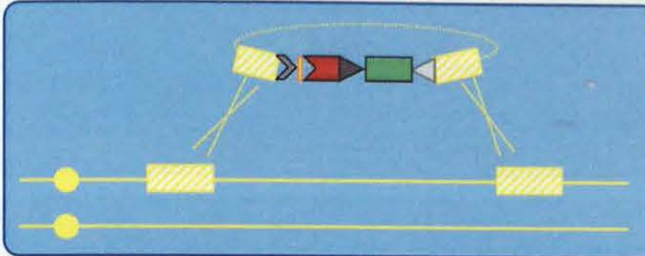
To reduce the time taken to generate an ES cell line with a *floxed* genomic interval, an alternative strategy that would *flux* the target locus with *lox* sites implements the plasmid-based gene-targeting construct, as shown in Fig. 7.2. The advantage of using such a construct is two-fold: first, it reduces the number of manipulations made to the ES cells; second, it removes the need to test whether the *lox* sites have been targeted *in cis*. The resultant ES cell line would harbour a heterozygous deletion for the locus of interest, which would limit this approach to loci that do not contain sequences that are hemizygous lethal.

The efficiency of RMCE that delivers sequences to the resultant *floxed*-deleted locus may be greater than for RMCE that exchanges intervals of equivalent size. The plasmid to deliver both *loxP* and *lox511* to the ES cell alpha-globin locus has been constructed and the RMCE efficiencies are to be assessed.

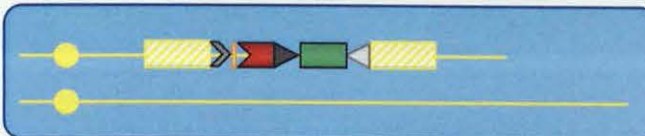
Fig. 7.2 Creating a floxed ES cell interval by a single gene-targeting event



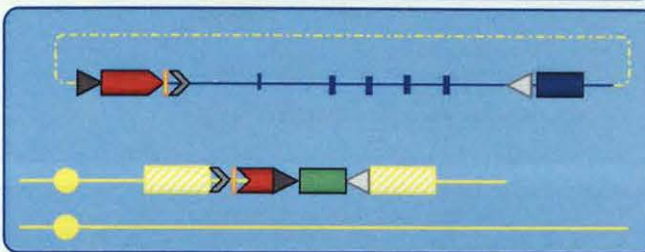
The locus of interest in the ES cell genome and position of the regions of sequence homology used for targeting.



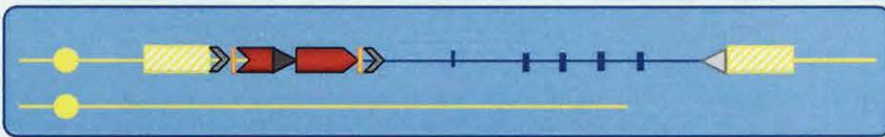
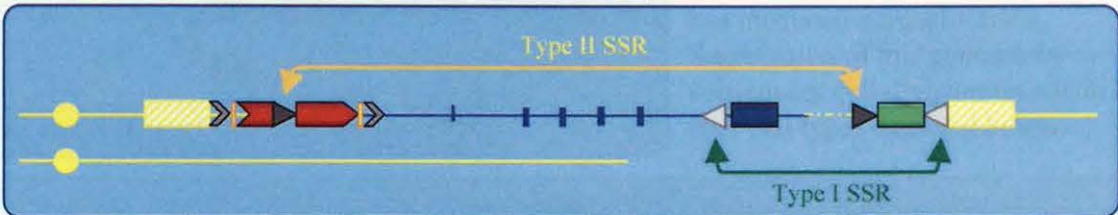
The neomycin marker, *flxed* with *loxP* and *lox511* is targeted, along with the 5'HPRT minigene region, by homologous recombination.



ES cell targeting results in neomycin resistant clones and creates a heterozygous deletion at the locus of interest.



SSR via *loxP* sites integrate the BAC molecule and generates the co-integrand structure, as depicted below. Type I SSR deletes a smaller interval than type II SSR, to produce the desirable RMCE event.

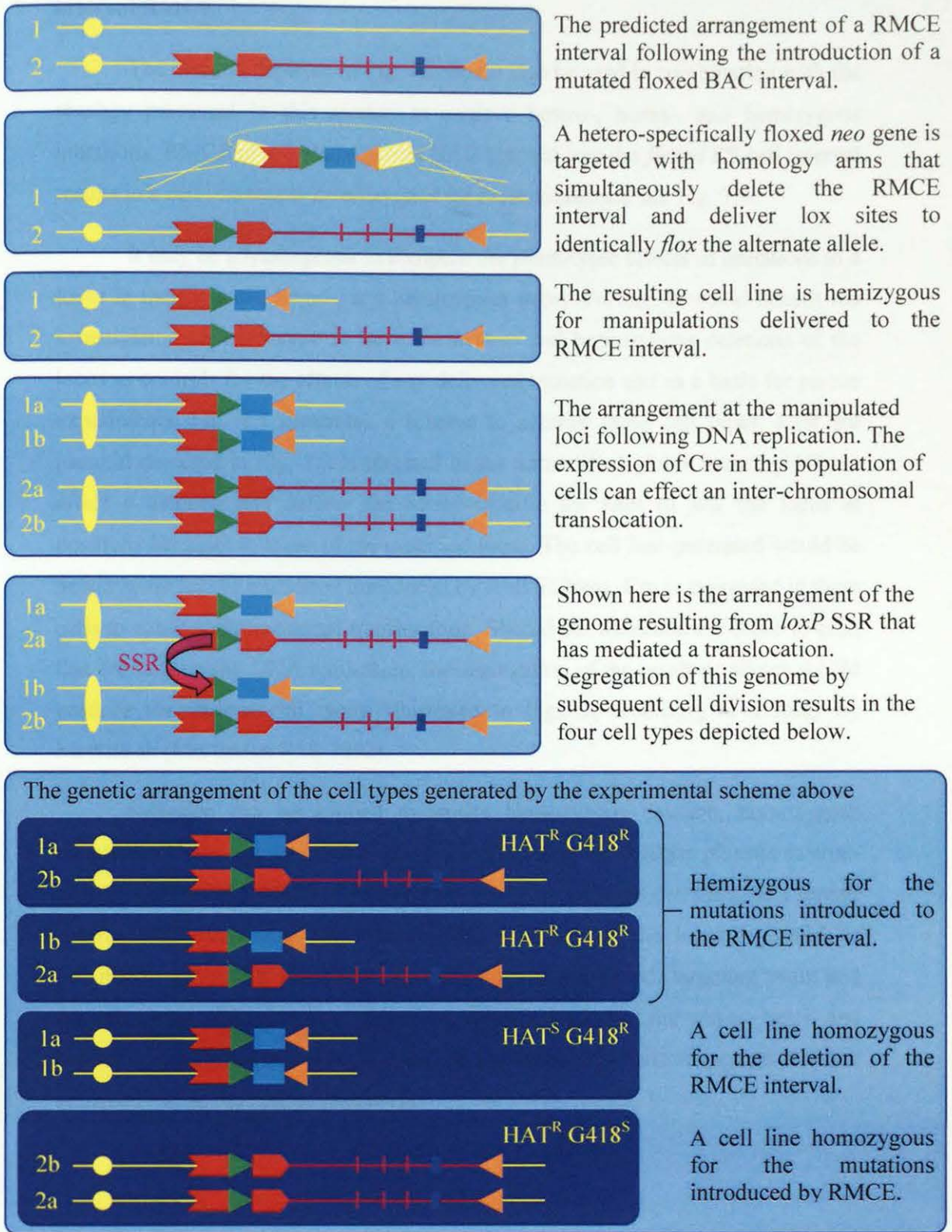


SSR between the BAC and ES cell locus results in RMCE.

Key

To overcome the need for two sequential targeting events to establish the *flxed* interval in the ES cell line, it would be possible to target the plasmid, as shown in the diagram, to the locus of interest by homologous recombination. The plasmid simultaneously delivers both hetero-specific *lox* sites and deletes the target locus. The resultant cell line is a heterozygous deletion line for the locus of interest, which may limit this approach to loci that do not contain sequences that are hemizygous lethal in ES cells.

Fig. 7.3 Generating homozygous and hemizygous loci following RMCE



7.4 Generating hetero-, homo- and hemizygous mutated loci following large interval RMCE

The plasmid depicted in Fig. 7.2, could also be used in combination with the strategy presented in this section to achieve hetero-, homo-, and hemizygous mutations. RMCE of a mutated *floxed* BAC interval into the *floxed* ES cell interval results in a cell line that is heterozygous for those mutations, see Fig. 7.3.

It may be advantageous to compare the phenotypic effects of mutations to a locus in their hetero-, homo-, and hemizygous states. Further, it would benefit the investigation on phenotype to have the hetero- and homozygous deletions of the locus as controls for the effects of any delivered mutation and as a basis for rescue experiments. Fig. 7.3 illustrates a scheme to achieve these cell types. First the plasmid depicted in Fig. 7.2 is targeted to the unmodified globin locus by HR, to effect a deletion and deliver the hetero-specific *lox* sites to *flux* the locus at positions identical to those of the modified locus. The cell line generated would be hemizygous for the mutations introduced by RMCE. Next, Cre is expressed in these cells to cause a chromosomal translocation. Should the translocation occur in cells that are undergoing DNA replication, the segregation of the resultant alleles should produce the various cell types, illustrated in Fig.7.3, according to a study by Mortensen (Mortensen RM, 1992).

Selection can be applied to isolate homozygous deletion, homozygous mutated and hemizygous mutated cell lines. Targeting the deletion plasmid to wild-type ES cells would therefore generate heterozygous deletion cell lines, as a further control. This illustrates that following manipulation of the locus it would be possible to generate a variety of cell types by just one further targeting event and the expression of Cre. Further, following characterisation of the homo-, hemi- and heterozygous phenotypes RMCE could be implemented to test sequences for their ability to rescue the mutant phenotype.

7.5 Mediating large interval RMCE in somatic cells

This strategy relies on gene-targeting in ES cells to establish the exchange interval however this should not restrict the application of this strategy to organisms from which ES cells have been derived. A number of somatic cell lineages have been demonstrated to support gene-targeting (as reviewed by Sedivy and Dutriaux, 1999), although due to the selection strategy implemented here the ablation of *hprt* function would be pre-requisite to the RMCE event. The combination of somatic cell targeting and nuclear transfer technologies (Campbell et al., 1996) would theoretically extend the application of this strategy to any organism, which would be most advantageous in the fields of xenotransplantation, bioreactor organism generation and orthologous sequence exchange investigations, as discussed later.

In addition to the availability of methods to generate *in vivo* models, the advantage of establishing the *floxed* interval in ES cells is that they may be differentiated into variety of cell types. Most somatic cell types do not support targeting by homologous recombination with efficiency, which makes the genetic manipulation problematic. However, deriving somatic cell lines from GMOs, in which the *floxed* interval has been engineered, would allow their genetic manipulation by RMCE. This would be of particular advantage to investigate cell types that do not support homologous recombination with efficiency or cell types that can not be obtained by ES cell differentiation.

The generation of mice by nuclear transfer of murine ES cell nuclei has been demonstrated (Wakayama et al., 1999). This development has the potential to significantly decrease the time necessary to generate murine models from genetically-engineered ES cells compared with the conventional route involving blastocyst contribution. This is because the step to generate a germline chimera is eradicated and furthermore the breeding programme required to generate the desired mouse model is either eradicated or reduced.

Another route that could decrease the time taken to generate a variety of murine models, each harbouring a distinct manipulation, would be to introduce the BAC and Cre recombinase at the one-cell embryo stage. This route would require

that the one-cell embryos be competent to mediate a RMCE event, which could be achieved by taking embryos from a mouse line derived from ES cells engineered to harbour the desired RMCE interval. In order to generate a variety of murine lines, each one-cell embryo would be micro-injected with an individual BAC clone from a pool of distinctly manipulated BACs, and Cre recombinase, to mediate the RMCE event. However the efficiency of the RMCE event, compounded with the efficiency of micro-injection, may make such a strategy impractical.

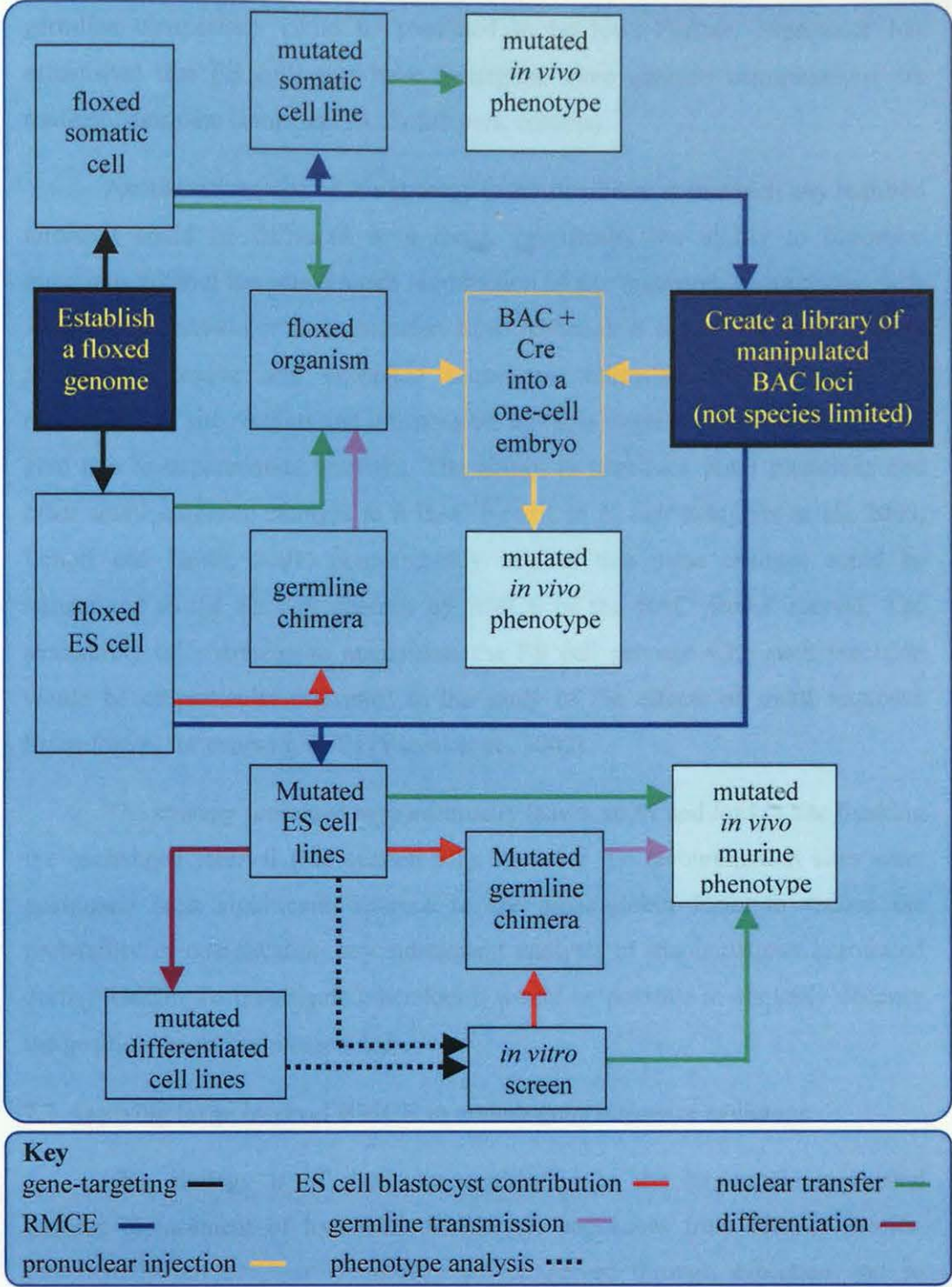
The methods available to manifest the RMCE *in vivo* and *in vitro* are summarised in Fig. 7.4.

7.6 Applying large interval RMCE to functional genomics

The strategy implemented here requires two gene-targeting events in ES cells to establish the exchange interval and a further Cre-*lox* mediated RMCE to introduce any manipulation to the locus. It could be argued that the desired mutation may be more easily generated using a single targeting event and since experience with ES cells has shown that increasing the number of manipulations may decrease germline capacity, it would therefore be counter-productive to perform more manipulations than necessary. However, it is envisioned that the strategy presented here will be applied to study loci that require a combination of manipulations to decipher function. As in many cases, the initial mutational analysis of a locus often leads to the requirement for further, more sophisticated genetic manipulations to better understand the function of a locus. The variety of manipulations that may be achieved is often time limited and compromised by the availability of technology to precisely mediate the desired manipulation. The potential of this strategy would be that a single *floxed* ES cell line could be generated and then used to produce many variously manipulated lines in a rapid manner, each by the single-step RMCE of BAC-harboured sequences, see Fig. 7.3.

The *E. coli* recombination system has been used to manipulate BACs in a rapid and versatile manner and would therefore facilitate the production of an unlimited collection of BACs harbouring the locus of interest with various mutational repertoires. Once the *floxed* ES cell line was established as germline

Fig. 7.4 Potential *in vitro* and *in vivo* manifestation of RMCE technology



The flow diagram illustrates the variety of methods available to produce *in vivo* and *in vitro* models harbouring the mutated genotypes created by RMCE.

competent, the likelihood of one further manipulation, that of RMCE, diminishing germline competency could be predicted to be low. Further, experience has established that ES cells that have undergone three genome manipulations are routinely germline competent (A. Smith pers. comms).

Another advantage of this strategy is the flexibility with which any required mutation could be delivered to a locus, specifically the ability to introduce mutations without the concomitant introduction of any undesirable sequences, such as selection markers or recombination sites. Although it is possible by *floxing* or *flirting* to remove any unwanted sequences, following SSR at least one recombination site remains and it is possible that this single recombination site may give rise to experimental artefacts. The ability to introduce point mutations and other small sequence changes to a BAC by HR in *E. coli* (Muyrers et al., 2000; Lalioti and Heath, 2001) concomitantly implies that these changes could be transmitted to the ES cell genome by RMCE of the BAC *floxed* interval. The availability of a strategy to manipulate the ES cell genome with such precision would be of particular relevance to the study of the effects of small sequence heterologies, for example SNPs (Vignal et al., 2002).

The strategy presented here eventually leaves an *frt* and *lox511* site flanking the exchanged interval (see section 5.1), however the recombination sites were positioned at a significant distance to the alpha-globin locus to reduce the probability of complicating any subsequent analysis of manipulations introduced during RMCE. To investigate other loci it would be possible to similarly distance the position the recombination sites.

7.7 Applying large interval RMCE to orthologous sequence exchange

This strategy is of particular application to the large-scale segmental genome replacement of loci with orthologous sequences from distinct species. Functionally advantageous sequences are conserved through evolution and so orthologous sequence exchange can be implemented to reveal those sequences that have conserved function between species and functional significance. Highly

conserved sequences are indicative of sequences essential for viability and comparatively, sequences that have undergone diversification may indicate an evolution of function, functional redundancy or a recent loss of functional significance.

Orthologous sequence exchange, implementing a strategy equivalent to that outlined here, is currently under investigation using a BAC containing the human alpha-globin locus to replace the murine locus in ES cells (H. Wallace pers. comms., Centre for Genome Research, Edinburgh University, 2001). Essential to the successful mediation of any orthologous sequence exchange is the definition of the orthologous region. Using comparative sequence analysis it is possible to establish the position of hetero-specific *lox* sites that should mediate a RMCE of functionally equivalent regions. The outcome of orthologous exchange of the murine alpha-globin locus with the equivalent human sequence should therefore determine whether the exchanged region is in fact functionally equivalent and should also generate useful models for human disease and drug testing. In addition, Douboule's group is attempting to apply a similar RMCE strategy utilizing BACs to investigate the phenotypic effects of exchanging the murine *hox* gene cluster with its human equivalent (F. Gonzalez pers. comms., Mouse Molecular Genetics conference, 2001).

Another potential application of orthologous sequence exchange is to optimise the production of organisms for organ xenotransplantation. At present, inter-species donation is flawed due to a high incidence of organ rejection. Segmental genome replacement could be used to exchange those sequences in the donor responsible for rejection to either produce a universal donor or a donor tailored to the requirements of the recipient.

7.8 Applying large interval RMCE to optimise transgene expression

In higher eukaryotes the regulation of transcription is determined, to a large extent by chromatin structure. This becomes especially important in cases where the expression of foreign genetic material is studied or where transgenes are introduced for the efficient production of pharmaceutically relevant proteins.

Although progress has recently been reported (Lipps and Bode, 2001), the development of episomal vectors that replicate and segregate in synchrony with the host cell is still in its infancy. Therefore, if stable long-term expression is desired, standard transfection systems are implemented to integrate transgenes into the genome of a bioreactor organism. However, these methods were developed with little consideration to control either the position in the genome into which the transgene integrated. This resulted in the position of transgenic integrations being largely random and lead to unpredictable expression patterns due to chromosome position-dependent effects. Additionally, the genetic arrangement of the transgene at its integration site was commonly as concatenated multiple copies of the transgene, which again lead to unpredictable transgene expression.

In contrast, the strategy tested in this thesis was designed to enable the delivery transgenes to a pre-defined genomic location and in a copy-number controlled manner. It could therefore be of advantage to the field of bioreactors by improving the efficiency with which the transgene product is produced (Kolb and Siddell, 1997). Further, the elucidation of transcriptionally permissive locations within a genome would permit the design of gene-therapy vectors to target such locations.

7.9 Applying large interval RMCE to investigate *in cis* regulatory sequences

Another means to optimise the expression of transgenes and gene-therapy vectors is to include sequences that buffer the constructs against chromosome position effects. The regulatory region of the beta-globin locus has proved to be significant since it has been shown to confer position independence and copy number dependence when included within transgenic constructs and was therefore subsequently termed a locus control region (LCR) (Fraser et al., 1990; Grosveld et al., 1987). Recent experiments however have challenged the ability of beta-globin LCR sequences alone to mediate characteristic gene expression profiles from different genome locations (Guy et al., 1996; Li et al., 2001).

The alpha and beta globin gene clusters result from a duplication event (Dickerson and Geis, 1983) and have since undergone sequence diversification.

Although each locus has similarities, with respect to gene order and developmental expression, it is improbable that the *cis*-acting distal regulatory regions of each locus originate from the same duplication event. Comparative sequence analysis between the murine beta-globin LCR and alpha-globin MRE reveals numerous differences in the location and number of both HSSs and protein binding domains. In addition the alpha-globin MRE is located in the first intron of a gene, *Prox1*, whilst the region surrounding the beta-globin LCR is devoid of gene sequences. However both loci are capable of co-ordinated expression throughout development to produce the haemoglobin tetramer. Transgenic studies have shown that the MRE is not an LCR since it can not produce expression patterns characteristic of sequences under LCR influence (Bouhassira et al., 1997). The comparative analysis of alpha and beta globin loci therefore provides a useful system to determine functionally significant sequences that regulate gene expression.

The strategy presented here facilitates the functional comparison of regulatory regions of the alpha and beta-globin loci. Comparative sequence analysis between the loci shows a range of sequence diversity and in order to re-capitulate each of these variations in a model system by a combinatorial approach using conventional gene-targeting technology would be labour intensive, inefficient and may produce artefacts due to the introduction of undesirable sequences, such as markers or recombination sites, as discussed previously. To exemplify this approach, the strategy described here could be implemented to facilitate the directed mutagenesis of the MRE harboured in a BAC by introducing sequences thought to confer LCR characteristics by successive rounds of targeting in *E. coli* by HR. The manipulated BACs could then be individually introduced to the *floxed* ES cell line and a comparative *in vitro* screen of haematopoietically differentiated ES cells could be used to determine the effects of successive mutations.

The investigation of the mechanism by which distal *cis*-acting regulatory elements function should not only further gene expression research but could also be applied to improve the design of transgenes. For example, an understanding of *cis*-acting sequences could be implemented to generate transgenic constructs with

more robust expression profiles, a facet that could be applied to the generation of more efficient bioreactor organisms.

7.10 Development of the large interval RMCE strategy in order to recover the exchanged interval

RMCE is thought to be a reciprocal reaction and so it may be possible to use this strategy as a basis to develop methods for the recovery of the genomic interval that is exchanged in ES cells. The reciprocal exchange of the alpha-globin locus residing in ES cells into the BAC vector was not investigated here and at present no other study has investigated the reciprocal nature of RMCE with intervals greater than 3 kb. To recover the reciprocal BAC molecule it would be necessary to develop the BAC vector to promote its stability and replication in mammalian cells (Cocchia M, 2000), in order to assist its purification from an overwhelming background of genomic DNA. In addition, the inclusion of an affinity tag to the BAC vector backbone may further facilitate the isolation of the reciprocal RMCE molecule. The tag could be selected for during DNA purification, thereby enriching the reciprocal molecule's concentration above that of the host cells' genomic DNA (Gates et al., 1996).

The ability to recover the reciprocal DNA molecule would be particularly useful in the investigation of chromatin structure. At present, the investigation of chromatin structure at specific loci is hampered by an overwhelming amount of background data, which is generated due to the presently unavoidable analysis of total genomic DNA. Therefore the ability to isolate and recover discrete genomic intervals that potentially retain their chromatin profiles would benefit the study of chromatin structure, re-modelling and its effects on gene expression, by simplifying the analysis of the data generated.

Appendix I

Software, Materials and Methods

I.1 Software

Sequence analysis and collation was performed using MacVector (v. 6.5.3) (<http://www.oxmol.com/prods/macvector>). Probe sequence analysis was performed using BLAST search (<http://www.ncbi.nlm.nih.gov/blast/newblast.html>).

I.2 Chemicals

Unless otherwise stated all chemicals were of analytical grade and supplied by BDH Laboratory (<http://www.cnbi.com>) supplies or Sigma (<http://www.sigmaaldrich.com>).

I.3 Oligonucleotides

Synthetic oligonucleotides were supplied by Oswel (http://www.oswel.com/code/en/prod_geno_olig.htm), their sequences are recorded by number in Appendix II.

I.4 Embryonic stem cells, culture media and solutions

I.4.1 E14Tg2a ES cells

E14Tg2a ES cells are a feeder-independent and rely on an exogenous source of differentiation inhibitory activity, DIA (also known as leukocyte inhibitory factor, LIF). By karyotype the cells are male and are originally derived from the 129/Ola strain of mice. These cells were produced to be HPRT deficient (Hooper et al., 1987; Thompson et al., 1989).

I.4.2 Cell culture medium

The culture medium can be stored for up to 1 month at 4°C

10x Glasgow MEM/BHK12 (Gibco), stored at 4°C	40 ml
7.5 % sodium bicarbonate (Gibco), stored at 4°C	13.2 ml
1x MEM non-essential amino acids (Gibco), stored at 4°C	4 ml
200 mM glutamine, 100mM sodium pyruvate (Gibco), stored at -20°C	8 ml
0.1M 2-mercaptoethanol (Sigma), stored at 4°C for 1 month maximum	0.4 ml
foetal calf serum (Globepharm, Surrey), batch tested, stored at -20°C	40 ml
sterile, deionised water	340 ml
DIA/LIF	0.44 ml

I.4.3 Differentiation inhibitory activity (DIA)

DIA was prepared by D. Rout at the CGR by transient expression of murine or human DIA/LIF expression plasmids in COS-7 cells using the method described by Smith (Smith et al., 1988). Serial dilutions of the supernatant were tested on ES cells. A 100-fold higher dilution than the minimal dilution required to keep ES cells undifferentiated was typically used.

I.4.4 Foetal calf serum

Serum batches were tested for their ability to sustain the growth, differentiation and viability of ES cells grown at clonal density in the presence and absence of DIA.

I.4.5 Phosphate Buffered Saline

One PBS tablet (Dulbecco A; Oxoid BR014G) was dissolved in 100 mls of water to give a PBS solution pH 7.3 that was filter sterilized.

I.4.6 Trypsin solution

Dissolve 250 mg of trypsin (Difco) and 372 mg EDTA disodium salt (Sigma) in 1 litre of PBS. Add 10 ml of chicken serum (Flow Labs) and filter sterilize. Stored in 20 ml aliquots at -20°C .

I.4.7 1 % Gelatin

Add 1 g of gelatin (Sigma) to 100 ml dH_2O , autoclave and store in 20 ml aliquots at 4°C . For a working solution of approximately 0.1 %, melt a 20 ml aliquot of 1 % gelatin in a 37°C waterbath and add to 200 ml of PBS.

I.4.8 0.1 M 2-mercaptoethanol

A working solution of 0.1 M was prepared by adding 100 μl of 2-mercaptoethanol (Sigma) was added to 14.1 ml PBS. Stored for up to 1 month at 4°C .

I.4.9 Freezing Solution

A 2x freeze solution of 20 % DMSO diluted with culture medium was typically prepared and used to freeze ES cells.

I.4.10 Quench Medium

Quench medium was prepared by mixing an equal volume of FCS with culture medium.

I.5 ES cell culture

Methods for routine culture of ES cells are based on those described in Smith, *et al.* 1991. Briefly, ES cells were maintained in a humidified incubator (Heraeus) at 7.5 % CO_2 and 37°C . All ES cell manipulations were carried out in a laminar flow sterile hood (ICN, flow). All solutions were warmed in a 37°C water bath prior to use. Tissue culture grade flasks and dishes (Corning) were gelatin coated (0.1 % in PBS, Sigma) and then aspirated before use. ES cell manipulations methods vary depending on whether the cells were being grown in flasks and dishes or in 96 well plates, usually following clonal selection of ES cells. This section describes both methods.

I.5.1 Passage and expansion of ES cells

Careful maintenance of ES cells was crucial to maintain the germline capacity should the cells be required for further investigation in an *in vivo* system. Medium was changed every other day and cells were not allowed to grow past 100 % confluence. Cells were most commonly passaged at a confluence of 80-90 % coverage of the culture vessel surface. Cells were passaged according to the following method:

1. ES cells were passaged once they had almost reached confluence. First the medium was removed by aspiration and cells were rinsed twice with PBS. Each wash of PBS was of sufficient volume to cover the surface of the culture vessel. Cells were not left without liquid coverage for more than 1 to 2 minutes.
2. Enough trypsin to cover the cell surface area was added (approximately 1 ml / 25cm²) and the flask incubated at 37°C for 1 to 2 minutes. After this time the cells are inspected, the flask tapped gently and re-inspected to ensure the cells are completely dislodged from the culture vessel surface. The vessel was returned to the 37°C incubator if required but the cells were not left in trypsin for longer than 4 to 5 minutes.
3. Once totally dislodged cells were made into a single cell suspension by re-pipetting 2 or 3 times and an equivalent volume of media added to stop trypsinization.
4. Cells were passaged into a new culture vessel at ratios of at least 1:2 and at most 1:6, depending on experimental requirements, made with culture media. It is important to maintain an adequate ES cell density to prevent differentiation.

I.5.2 96 well ES cell passage

1. 96 well plates were prepared by gelatinisation and the addition of 150 µl of culture media.
2. The 96 well plate to be passaged was washed twice with 100 µl of PBS using a multi-channel pipette.

3. 30 μ l of trypsin was added to each well and the plate placed in the 37°C incubator for 3 minutes.

4. Cells are pipetted and their adherence to the culture vessel observed. Once most of the cells were as a single cell suspension 70 μ l of quench media (see section I.4.10) was added.

5. 50 μ l of cells are placed in the prepared 96 well plate and to the remaining 50 μ l, 150 μ l of culture medium is added, to produce a 1:2 passage.

I.5.3 Freezing ES cells

The cryoprotector dimethyl sulphoxide (DMSO) diminishes the lethal effects of ice-crystals formation within the cell cytoplasm upon freezing but is however an inducer of ES cell differentiation and so exposure time was minimized before freezing and during thawing.

I.5.4 Freezing from a small flask

1. Cells were harvested as in steps 1 to 3, section I.5.1. Then transferred into a 20 ml Universal with 4ml GMEM and were pelleted by centrifugation at 1200 rpm for 5 minutes at room temperature (Denley BS400 Benchtop centrifuge).

2. Aspirate off media and add 1ml GMEM and 1ml 2x freezing solution.

3. Transfer to 2 cryotubes in the -70°C freezer.

I.5.5 Freezing a 96-well plate

The procedure 96 well plate freezing is identical to ES cell passage as described in section I.5.2, except for step 5. To freeze cells the original plate had 150 μ l of 2x freezing mix (section I.4.9) added instead of culture medium. The plates were taped and labelled and stored at -80°C.

I.5.6 Thawing vials of ES cells

- 1) A culture vessel with the required surface area for the number of cells being thawed was gelatinized, media added and placed in the 37°C incubator.

- 2) Cells were retrieved from liquid nitrogen storage and kept on dry ice until the moment to thaw.
- 3) The vial was held in the 37°C waterbath until the liquid begins to thaw and then the cells were transferred to the culture vessel with a plugged pasteur pipette.
- 4) The thawed cells were cultured for 4 to 5 hours to allow the cells to adhere to the vessel surface and then the culture medium was changed to reduce the concentration of DMSO.

I.5.7 96 well ES cell thawing

- 1) Warm the plate between hands.
- 2) Top up the wells with GMEM.
- 3) Aspirate off media, leaving 2mm (since cells can lift off).
- 4) Top up the wells with culture media.
- 5) Incubate until cells are confluent.

I.5.8 Electroporation of ES cells

Plasmid and BAC DNA was introduced into ES cells by electroporation. Cells and DNA are mixed in a cuvette and subjected to a high voltage electrical discharge. This is thought to lead to the formation of membrane pores through which DNA enters into the cell.

(a) Vector preparation for plasmid DNA

Plasmid DNA was prepared as described in section 1.6.1 (c). Electroporations typically introduced between 50 µg and 150 µg of DNA. Targeting constructs were linearised by restriction digest, phenol/chloroform extracted and precipitated with a 1/10th volume of sodium acetate and 2.5x volume of ice-cold ethanol. However, in some instances supercoiled plasmid DNA was electroporated and therefore did not require phenol/chloroform extraction before precipitation. The DNA was pelleted in an eppendorf by centrifugation at 13000 rpm (Heraeus,

Biofuge Pico centrifuge) for 10 minutes at 4°C. 1 ml of 70 % ethanol was used to rinse the inside of the eppendorf, taking care not to disturb the pellet, and then the eppendorf taken into the laminar flow hood. The 70 % ethanol was removed by pipette and the pellet left to air-dry in the hood. The DNA was thoroughly resuspended in 100 µl of PBS.

(b) Vector preparation for BAC DNA

BAC DNA for electroporation into ES cells was prepared as described in section I.6.1 (e). Between 50 µg and 150 µg of BAC DNA was electroporated into ES cells as supercoiled DNA. The DNA was prepared in an identical way to that of targeting vectors with respect to precipitation, the 70 % ethanol wash and resuspension in 100 µl of PBS. However when pelleting the DNA it was centrifuged at 10000 rpm (Eppendorf, Centrifuge 5415C) for 10 minutes at 4°C.

(c) Electroporation of DNA

- 1) Unless otherwise stated 1×10^8 cells were required for each electroporation. This number of cells is most easily harvested from 2 or 3 150 cm² culture flasks. Prior to the electroporation cells were cultured to ensure 70 to 80 % confluence on the day of electroporation. 3 to 4 hours before commencing the electroporation the culture media was replaced with 40 ml of fresh media per 150 cm² flask.
- 2) Two 150 cm² flasks were prepared by gelatinization, adding 35 mls of culture media and warmed by placing them in the 37°C incubator
- 3) Cells are washed and trypsinised as described in 2.5.2, steps 1 to 3.
- 4) After transfer into a 20 ml Universal they were pelleted by centrifugation at 1200 rpm for 5 minutes at room temperature (Denley BS400 Benchtop).
- 5) Cells were pooled and resuspended in a total of 10 mls of PBS, then counted
- 6) The volume of cells required for 1×10^8 cells was measured into a new 20 ml universal and re-pelleted, as described in step 3.

- 7) The pellet of cells was resuspended with PBS to give a total volume of 600 μl .
- 8) 100 μl of DNA (see 2.5.8 (a) and (b)) was added to the cells and carefully mixed by re-pipetting, ensuring no air bubbles were introduced into the mixture.
- 9) The cells and DNA were transferred to a 0.4 cm electroporation cuvette and capped. The cuvette was then pulsed in a Bio-Rad Gene Pulsar unit set at 3 μF , 0.8 kV. The time constant was noted and should be 0.1 msec.
- 10) Cells are allowed to recover for 10 minutes in the cuvette and then transferred carefully to a 20 ml universal. To ensure all the cells are transferred the cuvette is washed 2 or 3 times with 500 μl of culture media. The cells in the universal are made up to 4 mls with culture media and then split in half between the two 150 cm^2 flasks prepared in step 2.
- 11) Cells were cultured for 24 hours before clonal selection.

(d) Clonal selection of ES cells using selective media

- 1) An appropriate number of 10 cm diameter culture dishes were prepared by gelatinization, adding 8 mls of culture media and placing the dishes in the 37°C incubator.
- 2) Electroporated cells were inspected for the extent of cell death
- 3) Cells are washed and trypsinised as described in 2.5.2, steps 1 to 3.
- 4) After transfer into a 20 ml Universal cells were pelleted by centrifugation at 1200 rpm for 5 minutes at room temperature (Denley BS400 Benchtop).
- 5) Cells were pooled and resuspended in a total of 10 mls of culture media and then live cells were counted.
- 6) The volume of the cell suspension was adjusted to obtain a suspension of 5×10^6 cells. ml^{-1} , re-pelleting the cells if necessary.

- 7) Serial cell volume dilutions were made with culture media, e.g. 5×10^6 cells.ml⁻¹, 1×10^6 cells.ml⁻¹, 5×10^5 cells.ml⁻¹ and 1×10^5 cells.ml⁻¹. 1 ml aliquots for each dilution were then added to the prepared 10 cm diameter culture dishes.
- 8) 1 ml of a 10x stock of selective media was added to the selective plates. In addition control plates with no selection were maintained to enable a comparison for cell death to be made. Table 2.1 shows the selections used.

ES cell selections and killing rates

Selection	Effective concentration in ES cells	Days in selection to determine resistant clones
Hygromycin	120 µg.ml ⁻¹	12
G418	240 µg.ml ⁻¹	6
HAT	Hypoxanthine 0.1mM, Aminopterin 0.4µM, Thymidine 16µM	10
6-TG	1.67 µg.ml ⁻¹ (10mM)	4
Ganciclovir (GANC)	2 mM (MW 363.3)	5
Blasticidin	5 µg.ml ⁻¹	6

(e) Picking ES cell colonies

Clonal selection of ES cells in selective media gives rise to visible, discrete colonies of cells that have arisen from the expansion of a single resistant cell. In order to determine the genetic basis of this resistance it was necessary to isolate the colonies from each other and culture the cells to obtain a DNA preparation for analysis. This was achieved by using a pipette to take up the cell colony and place it in a single well of a 96 well plate.

I.6 Molecular biology methods

The methods described in this section are based on those described in Sambrook *et al.*, 1989, unless otherwise stated.

I.6.1 DNA Isolation and Purification

Overnight cultures grown for plasmid or BAC preparation contain the relevant antibiotic addition to the LB broth to maintain and propagate the DNA within the *E. coli*.

(a) Qiagen plasmid miniprep

The Qiagen miniprep method was used to prepare the recombination plasmid from overnight cultures of *E. coli* harbouring both BAC and plasmid. This was done to ensure the plasmid had been retained before using the culture to inoculate broth to prepare electrocompetent cells. The protocol is outlined at <http://www.qiagen.com>. Unless otherwise stated centrifugation takes place at 13000 rpm in a Heraeus, Biofuge Pico centrifuge at room temperature.

- 1) 1 ml of an overnight culture was centrifuged for 5 minutes.
- 2) The pellet is resuspended in 250 µl in Qiagen Buffer P1.
- 3) 250 µl of Qiagen Buffer P2 was added and the eppendorf was inverted to mix the contents.
- 4) 350 µl of Qiagen Buffer N3 was added and the eppendorf was inverted to mix the contents before centrifuging for 10 minutes.
- 5) The supernatant was applied to the Qiagen column and centrifuged for 1 minute.
- 6) 750 µl of Qiagen Buffer PE was added to the column and centrifuges for 1 minute.
- 7) The flow-through was discarded and the column centrifuged for a further minute.

- 8) The DNA was eluted into a fresh eppendorf by adding 50 μ l of TE and centrifuging for 1 minute.

(b) Plasmid miniprep

- 1) A single bacterial colony from an agar plate was used to inoculate a 5 ml overnight culture.
- 2) 1.5 ml of culture was transferred to an eppendorf and centrifuged at 13000 rpm (Heraeus, Biofuge Pico centrifuge) for 5 minutes at room temperature.
- 3) The bacterial pellet was resuspended on ice in 50 μ l of GTE (25mM Tris HCl, pH 7.5 (20°C); 1 mM EDTA, pH 8.0; 50mM Glucose).
- 4) 100 μ l of 0.2 M NaOH/ 1 % SDS was added, the tube mixed by flicking and left at room temperature for 5 minutes.
- 5) 75 μ l of 3M sodium acetate/acetic acid was added and the tube mixed by shaking.
- 6) The bacterial lysate was centrifuged as in step 2 and the supernatant removed by pipette into a fresh eppendorf.
- 7) A phenol/ chloroform extraction was performed. 100 μ l of dH₂O saturated phenol (company order no) and 100 μ l chloroform were added, the eppendorf was vortexed and centrifuged as in step 2. The top aqueous layer was transferred by pipette to a fresh eppendorf.
- 8) DNA was precipitated by adding 20 μ l of salt and 500 μ l of ice-cold ethanol and centrifuging at 13000 b for 10 minutes at 4°C.
- 9) The DNA pellet was washed with 200 μ l of 70 % ethanol and centrifuged as in step 2.
- 10) After air-drying the DNA pellet it was resuspended in 50 μ l of TE (10 mM Tris HCl, pH 8.0 (20°C); 1 mM EDTA, pH 8.0) with RNase added to a concentration of 0.5 mg.ml⁻¹.

(c) Plasmid maxiprep

Large amounts of plasmid DNA for cloning or the preparation of ES cell and BAC targeting DNA were prepared by the following method.

- 1) A single bacterial colony from an agar plate was used to inoculate a 5 ml overnight culture.
- 2) The overnight culture was used to seed a 250 ml overnight culture.
- 3) Cells were harvested by centrifugation at 3000 rpm (Sorvall Instruments, Du Pont, RC5C centrifuge fitted with a GSA rotor) for 20 minutes at 4°C.
- 4) The bacterial pellet was resuspended on ice in 10 ml of GTE (see section I.6.1(b)) and transferred to a 50 ml Corning tube.
- 5) 20 ml of 0.2 M NaOH/ 1 % SDS was added, the tube invert mixed and left at room temperature for 5 minutes.
- 6) 15 ml of 3M sodium acetate pH 4.8 was added and the tube invert mixed.
- 7) The bacterial lysate was centrifuged as in step 3 and the supernatant removed by pipette and split into two corning tubes, 20 ml in each.
- 8) DNA was precipitated by adding 25 ml of ice-cold isopropanol to each tube, mixing the contents by inverting the tube several times and then centrifuging immediately at 3000 rpm (MSE, Centaur2 centrifuge) for 10 minutes at 4°C.
- 9) The resulting pellets were pooled in 2 ml of TE, 2.5 ml of ice-cold 5 M LiCl added and left at 4°C for 15 minutes before centrifuging as in step 8.
- 10) The clear supernatant was transferred into a fresh tube and the DNA precipitated with 2.5 volumes of ice-cold ethanol and centrifuging as in step 8.
- 11) After the resulting pellet was air-dried it was resuspended in 500 µl of TE and transferred to an eppendorf and 20 µl of RNase at 10 mg.ml⁻¹ added. The eppendorf was incubated at room temperature for 30 minutes.

- 12) 500 μ l of cold 1.6 M NaCl₂, 13 % PEG (6000) was added, the tube was vortexed and incubated on ice for 15 minutes before centrifuging at 13000 rpm (Heraeus, Biofuge Pico centrifuge) for 10 minutes at 4°C.
- 13) The NaCl₂/PEG solution was removed by pipette, the eppendorf centrifuged at 13000 rpm (Heraeus, Biofuge Pico centrifuge) for 1 minute at room temperature and the remaining NaCl₂/ PEG solution removed.
- 14) Two extractions with an equal volume of chloroform, followed by two extractions with an equal volume dH₂O saturated phenol, followed by a final extraction with an equal volume of chloroform were performed.
- 15) The DNA was finally precipitated with 50 μ l of 3 M sodium acetate pH 5.5 and 1000 μ l of ice-cold ethanol. The tube was gently inverted until the precipitating DNA became visible. The DNA was then pelleted by centrifugation at 13000 rpm (Heraeus, Biofuge Pico centrifuge) for 3 to 5 minutes at room temperature.
- 16) The DNA pellet was washed with 70 % ethanol, centrifuged at 13000 rpm (Heraeus, Biofuge Pico centrifuge) for 3 minutes at room temperature, air-dried and resuspended in TE, typically a volume of 500 μ l.

(d) BAC Quick Prep

This method is used to prepare small amounts of BAC DNA to be investigated by restriction analysis, typically following a BAC targeting experiment to screen for correct integration of the targeting construct. For restriction analysis the entire preparation must be used since the yield of BAC DNA is small. In addition to visualize the BAC DNA the *E. coli* must be first cured of the recombination plasmid (see section I.6.8 (e)). The method is identical to the plasmid miniprep (see section I.6.1 (b)) with the following exceptions: all centrifugations are performed at 10000 rpm and not 13000 rpm (Heraeus, Biofuge Pico centrifuge); the phenol/chloroform extraction is omitted; and the preparation is mixed gently by inversion and at no time vortexed to reduce the shearing of the large BAC DNA molecule.

(e) Magnesil BAC Prep

From 500 ml of culture the typical yield was between 50 to 100 µg of BAC DNA.

- 1) 500 ml of L Broth was inoculated and grown overnight with shaking at 33°C.
- 2) The cells were pelleted at 3,000 rpm (Sorvall Instruments, Du Pont, RC5C centrifuge fitted with a GSA rotor) for 10 minutes at room temperature and resuspended in 6.25 ml of Cell Resuspension Solution in a 50 ml Falcon tube.
- 3) 6.25 ml of Cell Lysis Solution was added and mixed by inversion 4 or 5 times.
- 4) The suspension was incubated until it cleared (approximately 5 minutes).
- 5) 8.75 ml of Neutralization Solution was added and mixed by inversion.
- 6) The bacterial lysate was centrifuged at 4,000 rpm (MSE, Centaur2 centrifuge) for 15 minutes at room temperature and the supernatant filtered through fish filter wool placed in a 25 ml syringe.
- 7) The Endotoxin Removal Resin was completely resuspended and 1.25 ml added to the filtered supernatant. The tubes were inverted to mix the contents periodically during a 15 minute incubation at room temperature.
- 8) The tube mixed for the last time and then placed in the Magnetic Separation Unit and allowed to clear. The supernatant was then poured into a fresh 50 ml falcon.
- 9) 5 ml of 5 M Guanidine Thiocyanate was added to the supernatant and the tube mixed by inversion.
- 10) 3.75 ml of MagneSil Paramagnetic Particles were added and the tube incubated at room temperature for 15 minutes.
- 11) The tube was placed in the Magnetic Separation Unit and allowed to clear, the supernatant was then discarded.

- 12) The particles were washed with 5 ml of 4/40 Wash Solution, magnetically separated and the supernatant discarded.
- 13) The particles were washed three times using the method described in step 12 using 10 ml of 80 % ethanol. After the final wash the particles were air-dried.
- 14) 6 ml of dH₂O, pre-heated to 65°C, was added to the dry particle, vortexed and then left to reach room temperature. The particles were then magnetically separated and the supernatant containing the DNA transferred to a 15 ml Falcon tube. The tube was centrifuged at 3,000 rpm (MSE, Centaur2 centrifuge) for 10 minutes at room temperature to pellet any particles that may have been transferred. The supernatant was transferred to a fresh 50 ml Falcon.
- 15) The DNA was precipitated and pelleted by adding 2.5 volumes of ethanol and 1/10th volume of 3M sodium acetate. The DNA was pelleted at 3,000 rpm (MSE, Centaur2 centrifuge) for 10 minutes at 4°C.
- 16) The DNA pellet was washed with 70 % of ethanol and resuspended in 100 µl of TE.

(f) PCR purification with Qiagen

Qiagen columns were used to remove nucleotides, enzymes and other PCR reagents before electroporating the DNA into *E. coli*.

- 1) Five volumes of Qiagen Buffer PB were added to one volume of the reaction and then placed onto the column
- 2) DNA was bound to the column by centrifuging at 13000 rpm (Heraeus, Biofuge Pico centrifuge) for 1 minute at room temperature.
- 3) 750 µl of Qiagen Buffer PE was added to the column and centrifuged for 1 minute at 13000 rpm (Heraeus, Biofuge Pico centrifuge). To remove all traces of buffer the column was centrifuged at 13000 rpm (Heraeus, Biofuge Pico centrifuge) for 1 minute at room temperature.

- 4) To elute the DNA 30 μ l of Qiagen Buffer EB was added to the column, it was allowed to stand for 1 minute and then centrifuged at 13000 rpm (Heraeus, Biofuge Pico centrifuge) for 1 minute at room temperature.

(g) Gel purification with Qiagen

- 1) The required DNA restriction fragment was excised from the agarose gel with a clean scalpel.
- 2) The gel slice was placed in an eppendorf and weighed. Three volumes of Qiagen Buffer QG were added to one volume of gel.
- 3) The gel slice was dissolved by incubating the eppendorf at 50°C and then one gel volume of isopropanol was added.
- 4) DNA was bound to the column by applying the sample and centrifuging at 13000 rpm (Heraeus, Biofuge Pico centrifuge) for 1 minute at room temperature.
- 5) 750 μ l of Qiagen Buffer PE was added to the column and centrifuged for 1 minute at 13000 rpm (Heraeus, Biofuge Pico centrifuge). To remove all traces of buffer the column was centrifuged at 13000 rpm (Heraeus, Biofuge Pico centrifuge) for 1 minute at room temperature.
- 6) To elute the DNA 50 μ l of Qiagen Buffer EB was added to the column, it was allowed to stand for 1 minute and then centrifuged at 13000 rpm (Heraeus, Biofuge Pico centrifuge) for 1 minute at room temperature.

(h) Gel purification with GeneClean

- 1) The required DNA restriction fragment was excised from the agarose gel with a clean scalpel.
- 2) Three volumes of sodium iodide solution were added and the agarose melted by incubating at 50°C.

- 3) The amount of Glassmilk to be added was determined according to the amount of DNA in the excised band, based upon the ability of 1 μ l of Glassmilk to bind between 1 to 2 μ g of DNA. DNA was bound to the Glassmilk by incubation at room temperature for 15 minutes.
- 4) The Glassmilk, bound with DNA was pelleted by centrifugation at 13000 rpm (Heraeus, Biofuge Pico centrifuge) for 10 seconds at room temperature. The pellet was washed by resuspending in 50 volumes of NEW Wash (50 % ethanol). This step was repeated twice more but after the final centrifugation the pellet was resuspended in 50 μ l of TE.
- 5) The Glassmilk was pelleted and the supernatant containing the DNA transferred to a fresh eppendorf. The pellet was resuspended once again in 50 μ l of TE and re-pelleted and the supernatant added to the first 50 μ l elution.

(i) Removal of nucleotides and proteins by spin column

ProbeQuantTMG-50 Micro Columns were supplied by Amersham Biosciences (http://www4.amershambiosciences.com/APTRIX/upp01077.nsf/content/uk_homepage), product number 27-5335-01.

I.6.2 Enzyme reactions

(a) Plasmid restriction endonuclease digestion

Restriction endonucleases were supplied by either Roche or NEB BioLabs (<http://www.neb.com>). *I Sce I* is supplied by Roche (<http://www.roche-applied-science.com/>). Modifying enzymes were supplied by NEB BioLabs.

(b) ES cell genomic restriction endonuclease digestion

Spermidine was added to all genomic DNA restriction digests except those that required the addition of BSA. 30mM spermidine was prepared as a 10x solution for this purpose.

(c) PFGE ES cell plug restriction endonuclease digestion

See section I.6.6 (b).

(d) Alkaline phosphatase

Alkaline phosphatase treatment was used to remove the terminal phosphate group from linear DNA cloning vectors thereby reducing the frequency of vector self re-ligation and increasing the frequency of the desired ligation product.

- 1) A volume of TE equal to the digest reaction volume was added to the restriction linearised vector.
- 2) 1 Unit of alkaline phosphatase was added per μg of DNA and incubated at 37°C for 15 minutes.
- 3) 1/100 volume of 0.5 M EDTA was added and the solution heated to 75°C for 10 minutes to inactivate the enzyme
- 4) The products were either gel purified or phenol chloroform extracted, quantified and then used for further cloning.

(e) Blunting by end-filling with Klenow

- 1) The reaction constituents were typically 12 μl of DNA in dH_2O , 2 μl of T4 polymerase buffer, 4 μl of 500 mM dNTPs and 2 μl of T4 DNA polymerase at $3,000 \text{ U.ml}^{-1}$.
- 2) The reaction was incubated at 12°C for 20 minutes and stopped by incubating at 75°C for 10 minutes thereafter.
- 3) 20 μl of TE was added and DNA was precipitated and resuspended in TE for further cloning steps.

(f) Oligonucleotide probe labeling

- 1) A 20 μ l reaction mix containing 10 pmoles of oligonucleotide, 2 μ l of 10x T4 PNK buffer, 2 μ l of T4 PNK at 10,000 U.ml⁻¹, 10 μ l of [³²P]- γ -ATP and dH₂O was prepared and incubated at 37°C for 45 minutes.
- 2) The reaction was stopped by incubating at 65°C for 10 minutes and then passed through a spin column (see section I.6.1 (i)) by centrifuging at 3000 rpm (Heraeus, Biofuge Pico centrifuge) for 2 minutes at room temperature.
- 3) The probe was checked for activity using the Geiger counter and then placed in the hybridization solution.

(g) Rediprime probe labeling

Double stranded DNA probes produced by PCR or restriction were labelled using the 'rediprime II' random prime labelling kit (code RPN 1633) from Amersham Pharmacia Biotech (<http://www.apbiotech.com>) according to the manufacturers instructions.

- 1) 30 ng of probe DNA and 25 ng of control (λ Hind III restricted) DNA was resuspended separately in 45 μ l of TE.
- 2) The DNA was denatured by boiling in a waterbath for 5 minutes.
- 3) The DNA was snap-cooled by placing on ice for 5 minutes and then centrifuged to gather the tube contents.
- 4) Each DNA was added to a rediprime dehydrated reaction pellet and resuspended by pipette.
- 5) To the probe 5 μ l of radio-labelled [³²P]dCTP was added and to the control DNA 1 μ l was added. In addition 4 μ l of 0.5 M dCTP was added to the control DNA.
- 6) The tubes were flick mixed, centrifuged and then incubated at 37°C for 10 minutes.

- 7) The reactions were stopped by passing them through spin columns (see section I.6.1 (i)), centrifuging for at 3000 rpm (Eppendorf Centrifuge, 5415C) 2 minutes at room temperature.
- 8) The probes were briefly counted with a Geiger counter to determine their activity and then boiled in a waterbath for 5 minutes before being added to the hybridization solution.

(h) T4 DNA Ligase

Ligation reactions were performed using T4 DNA Ligase (NEB). The vector to insert molarities were typically 1:3 and 1:5. Controls of vector alone and insert alone reactions were conducted in parallel when possible to give an indication of the frequency of the desired ligation reaction. The reaction volume was typically between 20 and 30 μl with no more than a $1/10^{\text{th}}$ of the volume constituted by the T4 DNA ligase. Other reaction constituents were: T4 DNA ligase buffer, diluted from a 10x stock; vector DNA; insert DNA; dH_2O to the desired reaction volume. The reaction constituents were chilled to 4°C before combining the reagents on ice. All ligations were performed at 14°C overnight or for a minimum of 16 hours. The reaction products were determined following transformation (see section I.6.8 (b)) of the ligation into *E. coli*. Typically 10 μl of the ligation reaction was used for each 50 μl of competent cells.

I.6.3 Annealing oligos

30 pmoles of each oligonucleotide were used in a 20 μl reaction containing 2 μl of NEB Buffer 2 and dH_2O to make the reaction up to the required volume. The thin walled 500 μl PCR tube was placed into the PTC-200 Peltier Thermal Cycler and the annealing programme executed. The annealing programme heats the reaction to 100°C and controls the cooling by -1°C per 5 seconds until 24°C is reached. The PCR machine then maintains the annealed oligos at 4°C until required for the ligation reaction. The reaction contents were serially diluted 1:10 down to 10^{-3} of the original reaction's concentration and each dilution was used for ligation.

I.6.4 Quantification

(a) Optical Density

The concentration of a DNA solution was determined by optical density measurements. DNA was diluted to a concentration that produces an optical density reading at 260 nm between 0.1 and 1.0. Since 50 $\mu\text{g}.\text{ml}^{-1}$ of pure DNA gives an optical density at 260 nm of 1.0 the concentration of the DNA solution was derived from the following equation, where 'a' is the optical density reading and 'b' is the concentration of DNA in $\text{ng}.\mu\text{l}^{-1}$:

$$b = 50 \times \text{dilution of DNA sample} \times a$$

In addition the optical density ratio of the 260 nm and 280 nm reading was calculated with a figure above 1.7 indicating a pure DNA solution.

(b) Gel Quantification

To quantify smaller amounts of DNA (10 to 500 μg) a sample was run on a gel alongside known amounts of lambda Hind III restricted marker DNA at 100 $\text{ng}.\mu\text{l}^{-1}$. The band intensity of the known volume of sample was compared with the marker to find a band of similar size and intensity. Since the volume and concentration of marker loaded was known the amount of DNA in the band could be determined (see equation below) and hence the amount of DNA in the sample band determined. Once the amount of DNA in the sample band was known, the volume loaded and any dilutions that were made were considered to obtain the sample concentration.

$$(\text{Band size in kb}/48.5 \text{ kb}) \times 100 \text{ ng}.\mu\text{l}^{-1} \times \text{Volume of marker loaded in } \mu\text{l} = \text{DNA in the band in ng}$$

I.6.5 Automated sequencing

Sequencing was performed by the Sequencing facility at the CGR, by David Kivlichan. Sequencing reactions were performed using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit from Applied Biosystems as

described by the manufacturer, in a reaction that is a modified version of the dideoxynucleoside chain termination procedure described by Sanger *et al* (Sanger et al., 1977). Each terminating ddNTP was covalently linked to a fluorescent dye and the reaction products were separated on a denaturing 5 % polyacrylamide gel, in an ABI 377 automatic sequencer. Following the gel run the samples were scanned by the computer to check for correct tracking before extracting sequence information. The DNA template was prepared by Qiagen kit method (2.6.1 (a)), Maxiprep method (2.6.1 (c)), Magnesil BAC DNA (2.6.1 (e)) or by Qiagen purification of miniprep DNA (2.6.1 (a)).

I.6.6 Pulsed Field Gel Electrophoresis

(a) Making Plugs

- 1) Require 5×10^7 cells in each 1ml of agarose plugs, therefore grow up 3 x 150cm² flasks of cells, giving approx. 7.5×10^7 cells.
- 2) On the day to harvest the cells by trypsinizing set a waterbath to 50°C and melt the 2% agarose gently in a boiling waterbath. Once the agarose is melted equilibrate the solution to 50°C and leave whilst trypsinizing cells. In addition place an appropriate amount of cell suspension buffer in the waterbath.
- 3) Label eppendorfs and keep on ice ready to transfer cells to once trypsinized.
- 4) Trypsinize cells, pellet, resuspend in PBS, then pellet again, resuspend in 15 mls PBS, make a 1:10 dilution, then count.

$$\text{i) } \frac{\text{cell number} + \text{cell number}}{\text{(i) 2}} \times \text{dilution (1:10)} \times 10^4$$

- 5) Should have approx. 0.5×10^6 cells.ml⁻¹, so take 10 mls and pellet to obtain 5×10^7 cells. Resuspend the pellet in 1 ml of PBS, transfer to an eppendorf and keep on ice.
- 6) Pellet the cells by spinning at 1k 4°C for 5' (in cold room).

- 7) Resuspend the cells in an appropriate amount of Cell Suspension Buffer and put at 50°C for 5'.
- 8) For a 6.5 % agarose/cells use 690 µl of Cell Suspension Buffer and 310ul of 2% agarose.
- 9) Add the agarose to the cell suspension using a pipette and mix thoroughly by re-pipetting.
- 10) Transfer to the setting moulds (BioRad 170-3713) and place in the fridge (4°C) for at least 20 minutes (5 x 200ul into each well).
- 11) For each line into a 5ml culture tube add: 2.5ml Proteinase K reaction buffer; 100ul of Proteinase K; 1ml total of plugs. Incubate plugs in Proteinase K o/n at 50°C w/o agitation (can be left longer-up to 4 days).
- 12) To wash the plugs free of Proteinase K wash with 1xWash Buffer (make a 1:10 dilution of stock) with gentle agitation at room temperature. For each plug use 1 ml of wash and perform 4 x 1hr washes. Therefore for 5 x 200ul plugs use 5mls of 1x wash per hour per line, making a total of 20 mls of 1x wash per cell line.
- 13) For the second wash add 50 µl of 100mM PMSF to 5 mls 1x wash (1:100 to give a final concentration of 1 mM).
- 14) The plugs can be stored at 4°C for 3 months in 1x wash buffer.

(b) Digesting plugs

- 1) Place one plug (200µl) into a 1.5 ml eppendorf and wash for an hour in 1ml of 0.1x Wash Buffer.
- 2) After one hour decant off the wash buffer and add enough 0.1x wash buffer to cover the plug (for 200µl plug use 200µl 0.1x wash buffer).
- 3) Aspirate off this wash and add 1 ml of diluted 10x reaction buffer: 100µl buffer + 900 µl dH₂O. Leave for 1 hr with gentle agitation at room temperature.

- 4) Aspirate off the reaction buffer and add 200µl of 1x reaction buffer + 100 U of enzyme and incubate overnight at the required temperature.
- 5) Next day, wash the plug in 1x Wash Buffer for about 30 minutes with agitation.
- 6) Then, whilst waiting to load the gel place the plug into some running buffer.
- 7) Load about half of the plug into the gel.

(c) Loading plugs and running gel

Plugs were loaded into the gel slots and run according to the handbook guidelines for size separation for the BioRad CHEF Mapper PFGE apparatus (170-3654, 170-3655 and 170-3688).

I.6.7 Nucleic Acid Hybridisation and Screening

(a) Colony Hybridisation

Screening large numbers of colonies was preformed by colony hybridization. A 9 cm diameter, nylon membrane (Schliecher and Schuell; code 77500), was placed on top of the bacterial colonies. To orientate the filter it is first marked with a pattern of crosses using a pencil, then once placed on the plate the location of the crosses was marked on the plate with a permanent marker. Using forceps the filter was placed colony side up onto 3MM paper soaked in 5 % SDS / 2x SSC for 2 minutes. The filters were placed on dry 3MM paper and microwaved for between 30 seconds and two minutes until the colonies appeared as shiny discs. The filters were stored between 3MM paper until hybridized prior to which they were wetted in 2x SSC before being placed in the hybridization dish.

Colony hybridization screening was used to determine the inclusion of annealed oligonucleotides in a cloning reaction. The probe was typically one of the oligonucleotides and so the hybridization conditions were adjusted according to the nature of the oligonucleotide probe. The conditions of hybridization are shown overleaf.

2x Hybridisation Buffer

0.05 % sodium pyrophosphate

10x SSC

10x Denhardts

100 mg.ml⁻¹ yeast tRNA

The hybridization solution was diluted to a working solution with formamide and dH₂O. The percentage of formamide was adjusted depending on the length of the oligonucleotide probe.

The hybridization solution was pre-warmed to 42°C before adding the radio-labelled probe. The filters were hybridized overnight then washed and exposed to photographic film. The film was orientated to the filter using phosphofluorescent markers.

(b) Southern Blotting

Restricted DNA was electrophoretically separated and the gel denatured (1.5M NaCl₂; 0.5M NaOH for 30 minutes) and neutralized (1.5M NaCl₂; 0.5M Tris-HCl, pH 8.0 for 30 minutes) before the DNA was transferred to positively charged nylon membranes (Roche, product number 1-417-240).

Following blotting the filters were air-dried and the fixed using UV irradiation, 600 J at a wavelength of 400 nm. The filters were washed briefly in 2x SSC before being placed in pre-hybridisation conditions for at least 1 hour. The pre-hybridisation solution was then discarded and hybridization solution added to the filters along with radio-labelled probes. The filters were hybridized overnight at 65°C, washed, wrapped in Saranwrap and exposed to photographic film.

Genomic Southern blot filters are hybridized under the following conditions:

20x Denhardt's, stored at -20°C.

Ficoll 400	4 g.l ⁻¹
PVP	4 g.l ⁻¹

BSA Fraction V	4 g.l ⁻¹
----------------	---------------------

Hybridisation buffer, stored at 4 °C for up to 1 year.

Na ₃ Citrate	22.5 g.l ⁻¹
NaCl ₂	44.8 g.l ⁻¹
Dextran sulphate	85.2 g.l ⁻¹
20 % SDS	21.3 ml.l ⁻¹
20x Denhardt's	203 ml.l ⁻¹

The solution was warmed to 65°C and filtered through a 0.45 µm Nalgene filter.

I.6.8 Recombination mediated cloning

(a) Making CaCl₂ competent cells

250 ml of LB broth was inoculated with 5 ml of an overnight culture and cultured at 37°C with shaking until an optical density at A₆₀₀ of 0.5 was achieved. Two 25 ml aliquots of cell culture were pelleted by centrifugation at 4000 rpm (MSA, Centaur2 centrifuge) for 10 minutes at 4°C. The pellet was resuspended in 12.5 ml of ice-cold 50 mM CaCl₂ and incubated on ice for 15 minutes. Cells were again pelleted, 4000 rpm (MSA, Centaur2 centrifuge) for 10 minutes at 4°C, and resuspended in 2.5 mls of Freezing Mix. The pellets were pooled and 50 µl aliquots dispensed into eppendorfs on dry ice. The cells were stored at -70°C.

(b) Bacterial transformation

Chemically competent cells prepared as described in section I.6.8 (a) were used to introduce circular plasmid DNA to *E. coli*. The plasmid DNA was typically either the products of a ligation reaction or recombination plasmids used during the *in vivo* cloning reactions.

DNA was mixed with an aliquot of competent cells that had been thawed on ice. The DNA/cell mix was incubated on ice for 30 minutes on ice and then heat-shocked in a 42°C waterbath for 1 minute. 1 ml of LB broth was added and the cells were transferred to a loose top tube and then allowed to recover for 60 minutes in a 37°C incubator with shaking. Cells were then plated onto agar plates with the relevant antibiotic selection media included. Plates were incubated at 37°C, no

shaking, overnight. The colonies that arose were analysed to determine the nature of the plasmid DNA they harboured. To test a new batch of competent cells 10 pg of pBS(KS⁺) was used and the number of colonies arising on ampicillin plates from 10 µl of plated culture was recorded to give an indication of the cells' competency.

(c) Making electrocompetent cells

During the preparation of electrocompetent cells it was imperative that all reagents and materials coming into contact with the cells were kept as cool as possible. For this reason the following preparative steps were made:

- 1) Pipettes were placed in a -80°C freezer for at least an hour before their use.
- 2) The centrifuge SS-34 rotor was cooled to -4°C by spinning at 3000 rpm (Sorvall Instruments, Du Pont, RC5C centrifuge fitted with an SS-34 rotor) for 20 minutes before use.
- 3) The centrifuge tubes were chilled in ice for at least 2 hours before use.
- 4) All solutions were autoclaved and prior to use placed on ice for at least 2 hours before use to ensure their temperature was 0°C.

The cells were prepared by the following method:

- 1) Ten overnight cultures were set up in L-Broth and the relevant antibiotic selection in loose-capped tubes and incubated overnight at 37°C with shaking.
- 2) The next day 4 x 100 mls of L-broth were pre-warmed to 37°C in the incubator and 1 ml aliquots were taken of each overnight culture and DNA prepared by Qiagen miniprep method. The DNA was restricted and analysed to determine whether the cells were maintaining the recombination plasmid. One positive culture was selected and used to inoculate the pre-warmed L Broth with 1 ml of culture.
- 3) The cells were incubated at 37°C with shaking and their growth monitored by optical density measurements at 600 nm wavelength.

- 4) When an OD reading of 0.1 was obtained, a freshly made stock of 20 % arabinose was added to the culture to a final concentration of 0.2 %, to induce the activity of the genes on the recombination plasmid driven by the pBAD promoter. The minimum culture period resulting in maximal induction was 20 minutes.
- 5) When the cells had reached an OD of 0.4 (ideal) the cells were removed from the incubator and cooled on ice for 5 to 10 minutes.
- 6) 50 ml of cells were centrifuged in pre-chilled 75 ml centrifuge tubes at 6000 rpm (Sorvall Instruments, Du Pont, RC5C centrifuge fitted with an SS-34 rotor) for 8 minutes at -4°C in a pre-chilled SS-34 rotor. The supernatant was discarded and a further 50 ml added and the cells pelleted under the same conditions as before.
- 7) The supernatant was discarded and the cells were resuspended in 10 mls of 10 % glycerol chilled to 0°C by incubating on ice (minimum 1 hour) using pre-cooled pipettes. Before dispensing the 10 % glycerol the pre-chilled pipette has liquid drawn through it to further cool the pipette. The cells are thoroughly resuspended keeping them on ice and then centrifuged at 6000 rpm (Sorvall Instruments, Du Pont, RC5C centrifuge fitted with an SS-34 rotor) for 8 minutes at -4°C .
- 8) Step 7 was repeated until the cells were centrifuged for a total three times. The supernatant from the last centrifugation is poured away and the remaining liquid used to resuspend the cell pellet. The resulting cell volume was typically 500 μl for effective electrocompetent cells.
- 9) 50 ml aliquots of the cells were made into pre-chilled eppendorfs and used the same day for electroporation.

(d) Electroporating bacteria

Aliquots of electrocompetent cells were thawed on ice. The DNA was added and mixed thoroughly with the cells by pipetting and then transferred to pre-cooled

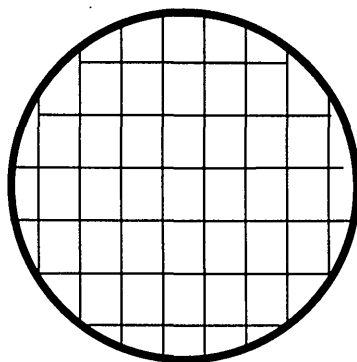
electroporation cuvettes (BioRad *E. coli* Pulser Cuvette, 0.2 cm gap, 165-2086). The cells were kept on ice, electroporated at 2.5 kV (BioRad Gene Pulser, 25 μ F with pulse controller set to 200 ohms), returned to ice for 10 minutes and then added to 1 ml of L-broth in a loose-capped tube. The cells were allowed to recover for 1 hour in a 37°C incubator with shaking and then plated with the relevant antibiotic selection.

(e) NiCl_2 curing of *E. coli* cells harbouring a tetracycline^R plasmid

- 1) A 5 ml overnight culture was grown from either a single colony on an agar plate or from a 100 μ l aliquot from a glycerol stock. The L-broth contains relevant antibiotics but not tetracycline. In addition 50 μ l of 500 mM NiCl_2 is added to give a final concentration of 5 mM, i.e. a 1:100 dilution.
- 2) 50 μ l of the overnight culture was plated onto agar with and without 5 mM NiCl_2 and the number of resulting colonies assessed to determine the percentage that had lost the tetracycline resistant recombination plasmid.
- 3) An adequate number of overnight cultures (without NiCl_2) were prepared from the agar plate containing the NiCl_2 to ensure that a culture without the tetracycline resistant recombination plasmid was obtained.
- 4) DNA from the overnight cultures was prepared by the Quick Prep method and analysed by restriction analysis to determine whether the recombination plasmid was lost and also the integrity of the BAC DNA.
- 5) Overnight cultures with cells harbouring the required DNA were used to prepare glycerol stocks, from which large preparations of DNA could be prepared using the Magnesil method.

(f) Replica plating

The colonies that resulted from electroporation were initially screened by replica plating. Agar plates were prepared with the relevant antibiotics and then identically grided as indicated in the diagram below.



Colonies arising from the electroporation plate were picked with a 20-200 μ l pipette tip and streaked onto the surface of successive agar plates, within the identical perimeter square for each. The order that the plates were streaked was significant and therefore maintained throughout the procedure. The last plate that was streaked contained chloramphenicol to ensure that any clone containing the BAC was maintained and as a positive control for the other plates. In this way, the antibiotic resistance profiles for each of the colonies were determined.

Table of antibiotic resistances for plasmids and BACs

Antibiotic	Stock concentration	Plasmid selective concentration	BAC selective concentration
Ampicillin	50 mg.ml ⁻¹ in dH ₂ O	50 μ g.ml ⁻¹	25 μ g.ml ⁻¹
Kanamycin	10 mg.ml ⁻¹ in dH ₂ O	10 μ g.ml ⁻¹	5 μ g.ml ⁻¹
Chloramphenicol	25 mg.ml ⁻¹ in ethanol	25 μ g.ml ⁻¹	12.5 μ g.ml ⁻¹
Blasticidin	10 mg.ml ⁻¹ in dH ₂ O	100 μ g.ml ⁻¹	50 μ g.ml ⁻¹
Tetracycline	6.25mg.ml ⁻¹ in ethanol	12.5 μ g.ml ⁻¹	Not determined

Appendix II

Oligonucleotide sequences

Oligo Number	Sequence (5' to 3')
A4128	GATCTTAGGGATAACAGGGTAAT
A4129	GATCATTACCCTGTTATCCCTAA
A4891	GATCCATAAAGTTCGTATAATGTATGCTATACGAAGTTAT
A4892	GATCATAAAGTTCGTATAGCATACATTATACGAAGTTATG
A5136	TCGAGGGATCCATAAAGTTCGTATAATGTATACTATACGAAGTTAT
A5137	TCGAATAAAGTTCGTATAGTATACATTATACGAAGTTATGGATCCC
B8141	GCCGAATTCCCTCGACCATGCCCAA
B8146	GCCGAATTCTCTCAGACTGTGGCAGG
C0367	ATTCAAGTAGCCTTCCCAGCAGTCACAGCAATAAAGTTCGTATAGCAT
C0368	AAAGTGTAGTCCCTCTCCTCTGGTGCCTGGATAAAGTTCGTATAATGT
C0465	GCCGGATCCGCATGCCTGCAGGTCTG
C0466	GCCGGATCCAGACATGATAAGA
D4329	CACTCCGCAGGAAGCAAC
D4330	AGAGGCAAGAGCTACAGAAAC
D4331	TCCACATTACATTGGTGAGA
D4332	CCAATGATGGATCCGTAAG
D6624	CCCGGTGACCCGGAAGCAGGGGCGGCAT
D6625	CCCGGGATCCGACAGCCACCTCCTGCC
D6626	TTAGCGGCCGCGCTATATCTTCATGATC
D6627	CCCGGGATCCAAGTCTTCTACACTGC
F10312	CCCGGTGACGAAATGGTCTTGTATATCCCAGG
F10313	TTAGCGGCCGCTCCTGCAGCATGAGGAGCG
F10314	TTAGCGGCCGCAAGCTTTGGTTAGCTTCTGAC
F10315	TTAGCGGCCGCGTTCAGCAAGCTCCAACAC
F10316	CCCGGGATCCTAAGTTGGCTCTCTAGTTCCA
F10317	CCCGGGATCCGAGCATTATCTGAAATTGCCTC
F10318	TTAAGGGATGGGACACTTG
F10319	CCATATGAGGAAGTTTGGA
F10320	AACTCAGTGATCCTACCTC
F10321	ACAATGGAACTCTGTCTAAG
F1346	TTTCTCAGTATGGGCAGG
F1347	TGGCTTCCTAGAGAATGG
F22433	GATCATAAAGTTCGTATAATGTATACTATACGAAGTTATCCATGGATCCG TAC
F22434	GGATCCATGGATAAAGTTCGTATAGTATACATTATACGAAGTTAT
F53524	CGCGCCCCCACCACACACACACACACTCAGGCAGTAAG GGCAGGAGGTGGCTGTGCGGATCCAAGTCTTCTACACTGCCAGACCTCC ATATATGTCAACCCTGCCTCTGCAGATACAACCAAGG
F53525	CGCGCCTTGGTTGTATCTGCAGAGGCAGGGTTGACATATATGGAGGTC

	TGGCAGTGTAGAAGACTTGGATCCGACAGCCACCTCCTGCCCTTACTG CCTGAGTGTGTGTGTGTGTGTGGGTGGGGGTGGGGGG
F53526	TAGCTCCAGCTCTGCACTAAGC
F53527	AAGGTTGGTTGTATCTGCAGAGG
T3	AATTAACCCTCACTAAAGGG
T7	GTAATACGACTCACTATAGGGC
W2186	GATCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCATTACCCTGT TATCCCTA
W2187	GATCTAGGGGATAACAGGGTAATGAAGTTCCTATACTTTCTAGAGAATA GGAACCTCG
Z6205	GATCCATAACTTCGTATAATGTATACTATACGAAGTTATAGATCTATAA CTTCGTATAATGTATACTATACGAAGTTATG
Z6207	GATCCATAACTTCGTATAGTATACATTATACGAAGTTATAGATCTATAA CTTCGTATAGTATACATTATACGAAGTTATG

Appendix III

Plasmids and BACs

Chapter 2 ES Cell Targeting

Number	Name
Plasmids used and generated in the construction of the ES cell downstream targeting construct: <i>lox511/hygromycin/I Sce 1/lox511</i>	
CP#50	pBS(KS+)
CP#20	pBS/511/Bgl II/511
AJHS#948	pHA58
CP#41a	pBS /511/ PGK-HYG /511
CP#41b	pBS /511/ SCE / PGK-HYG /511
CP#44	pBS/ DnLHA
CP#45	pBS/ DnLHA / DnRHA
CP#47	pBS/ LHA /511/ SCE / PGK-HYG /511/ RHA (3)
CP#48	pBS/ LHA /511/ HYG-PGK / SCE /511/ RHA (4)
Plasmids used and generated in the construction of the ES cell upstream targeting construct: <i>frt/I Sce 1/5'hpvt/loxP/neomycin</i>	
CP#1	pBS/ UpLHA
CP#7	pBS/ UpLHA / UpRHA
AJHS#1171	frt /Sce /5'HPRT/neo
CP#11	1171 (hpvt-neo) + Up/L/R

Chapter 3 Characterisation of the double-targeted ES cell lines

Number	Name
Plasmids used and generated in the construction of the test plasmid to delete the <i>loxP-lox511</i> interval in the CAT ES cell lines	
AJHS#924	<i>loxP/3'HPRT</i>
CP#51	924+FRT (51)
CP#98	924/FRT + 511 oligo
Plasmids used and generated in the construction of the Cre recombinase plasmid	
CP#93	pNPKCreERc (www.embl-heidelberg.de/ExternalInfo/stewart/ plasmids.html)
CP#15	Austin#575: pCAGSIP
CP#64	pCAGG.CreEBD.ires.puro
CP#99	pCAGG.Cre.ires.puro

Chapter 4 BAC Targeting

Number	Name
Plasmids used and generated in the construction of the BAC downstream targeting construct: <i>lox511/kanamycin/lox511/HSVtk</i>	
CP#140	PKaX
CP#141	pKaX (HIII to BHI)
CP#20	pBS/511/Bgl II/511
CP#24	pBS/511/kan(r&c)/511 (r&c)
AJHS#990	990/pBS-MC1tk/-3
CP#35	TK with A
CP#37	TK with A (<i>Xho</i> I to BHI)
CP#45	pBS/ DnLHA / DnRHA
CP#37 and 38?	pBS/LHA/kan-TK/RHA
CP#137	pBS (KS+) Asc I oligos
CP#139	pBS/ Asc /kan-TK/Asc
Plasmids used and generated in the construction of the BAC upstream targeting construct: <i>blasticidin/loxP/3'hprt/I Sce I/frt</i>	
CP#94	HN#381: mPGK pEM7 BSD pA
AJHS#1172	mPGK Hyg / 3'HPRT
CP#109	1172/94(BSD// <i>loxP</i> /3'HPRT/SceFRT)
CP#7	pBS/ UpLHA / UpRHA
CP#113	UpL/BSD-3'HPRT/UpR
The plasmid used to mediate the targeting by recombination mediated cloning in <i>E. coli</i>	
CP#114	pR6K / BAD / $\alpha\beta\gamma$ (tet ^R)
The BAC encompassing the alpha-globin locus and derived molecules following targeting	
CP#117	pBAC14567
CP#120	pBAC14567/TK
CP#111	pBAC14567/TK/3'HPRT
The BAC molecules recovered following the <i>in vitro</i> Cre analysis	
CP#125	pBAC14567/TK/ 3'HPRT ^{-BSD}
CP#128	pBAC14567/TK ^{-KAN} / 3'HPRT
CP#131	pBAC14567/TK ^{-KAN} / 3'HPRT ^{-BSD}

Chapter 5 Large interval RMCE

Number	Name
The BAC molecule used to mediate the large interval RMCE event	
CP#111	pBAC14567/TK/3'HPRT
The plasmid used to express Cre recombinase	
CP#99	pCAGGsCre.ires.puro

Appendix IV

PCR conditions

Summary

IV.a to d probes: UpLP, UpRP, DnLP, DnRP

IV.e to h ES cell targeting homology arms: UpLHA, UpRHA, DnLHA, DnRHA

IV.a PCR Conditions to make Upstream Left-hand Probe (UpLP)

Reaction Components:

Cosmid DNA 100 ng.μl ⁻¹	5 μl
Oligo F10321 3 pmol.μl ⁻¹	20 μl
Oligo F10318 3 pmol.μl ⁻¹	20 μl
dNTPs 1mM	40 μl
Taq Buffer including Mg 15mM	20 μl
dH ₂ O	90 μl
Mix and then centrifuge before adding Taq	
Taq polymerase	5 μl

Cycling parameters:

Step 1	Denature	95 °C	1.00
Step 2	Anneal	51 °C	0.30
Step 3	Elongate	72 °C	1.00
Step 4	Goto step 1	30 times	
Step 5	Hold	4°C	
Step 6	End		

IV.b Conditions to make Upstream Right-hand Probe (UpRP)

Reaction Components:

E14TG2a ES cell DNA 100 ng.μl ⁻¹	5 μl
Oligo F10319 3 pmol.μl ⁻¹	20 μl
Oligo F10320 3 pmol.μl ⁻¹	20 μl
dNTPs 1mM	40 μl
Taq Buffer including Mg 15mM	20 μl
dH ₂ O	90 μl
Mix and then centrifuge before adding Taq	
Taq polymerase	5 μl

Cycling parameters:

Step 1	Denature	95 °C	1.00
Step 2	Anneal	50 °C	0.30
Step 3	Elongate	72 °C	1.00
Step 4	Goto step 1	30 times	
Step 5	Hold	4°C	
Step 6	End		

IV.c Conditions to make Downstream Left-hand Probe (DnLP)

Reaction Components:

E14TG2a ES cell DNA 100 ng.µl ⁻¹	10 µl
Oligo D4331 from stock	8 µl
Oligo D4332 from stock	10 µl
dNTPs 1mM	40 µl
Taq Buffer including Mg 15mM	20 µl
dH ₂ O	110 µl
Mix and then centrifuge before adding Taq	
Taq polymerase	2.0 µl

Cycling parameters:

Step 1	Denature	98 °C	0.20
Step 2	Anneal	55 °C	0.30
Step 3	Elongate	72 °C	0.30
Step 4	Goto step 1	30 times	
Step 5	Elongate	72 °C	2.00
Step 6	Hold	4 °C	
Step 7	End		

IV.d Conditions to make Downstream Right-hand Probe (DnRP)

Reaction Components:

E14TG2a ES cell DNA 100 ng.µl ⁻¹	5 µl
Oligo D4329 from stock	8.0 µl
Oligo D4330 from stock	8.0 µl
dNTPs 1mM	40 µl
Taq Buffer including Mg 15mM	20 µl
dH ₂ O	90 µl
Mix and then centrifuge before adding Taq	
Taq polymerase	2.0 µl

Cycling parameters:

Step 1	Denature	9 °C	0.20
Step 2	Anneal	55 °C	0.30
Step 3	Elongate	72 °C	0.30
Step 4	Goto step 1	30 times	
Step 5	Elongate	72 °C	2.00
Step 6	Hold	4 °C	
Step 7	End		

IV.e Conditions to make Upstream Left Homology Arm (UpLHA)

Reaction Components:

E14Tg2a DNA 5 $\mu\text{g} \cdot \mu\text{l}^{-1}$	1 μl
Oligo F10315 50 $\text{ng} \cdot \mu\text{l}^{-1}$	5 μl
Oligo F10316 50 $\text{ng} \cdot \mu\text{l}^{-1}$	5 μl
dNTPs 1mM	20 μl
Pfu Buffer	10 μl
dH ₂ O	8 μl
Mix before adding Pfu	
Pfu polymerase	1 μl

Cycling parameters:

Step 1	Denature	98 °C	2.00
Step 2	Denature	98 °C	1.00
Step 3	Anneal	53 °C	1.00
Step 4	Elongate	72 °C	3.00
Step 5	Goto step 2	14 times	
Step 6	Elongate	72 °C	10.00
Step 7	Hold	4°C	
Step 8	End		

IV.f Conditions to make Upstream Right Homology Arm (UpRHA)

Reaction Components:

pBAC14567 DNA 100 ng.μl ⁻¹	10 μl
Oligo F10312 50 ng.μl ⁻¹	5 μl
Oligo F10317 50 ng.μl ⁻¹	5 μl
dNTPs 1mM	20 μl
Pfu Buffer	10 μl
dH ₂ O	49 μl
Mix before adding Pfu	
Pfu polymerase	1 μl

Cycling parameters:

Step 1	Denature	98 °C	2.00
Step 2	Denature	98 °C	1.00
Step 3	Anneal	56 °C	1.00
Step 4	Elongate	72 °C	3.00
Step 5	Goto step 2	14 times	
Step 6	Elongate	72 °C	10.00
Step 7	Hold	4°C	
Step 8	End		

IV.g Conditions to make Downstream Left Homology Arm (DnLHA)

Reaction Components:

pBAC14567 DNA 100 ng.µl ⁻¹	10 µl
Oligo D6624 x pmol.µl ⁻¹	2.5 µl
Oligo D6625 x pmol.µl ⁻¹	2.5 µl
dNTPs 2mM	20 µl
Pfu Buffer	10 µl
dH ₂ O	54 µl
Mix and then centrifuge before adding Taq	
Pfu polymerase	1 µl

Cycling parameters:

Step 1	Denature	94 °C	1.00
Step 2	Anneal	54 °C	1.00
Step 3	Elongate	72 °C	3.00
Step 4	Goto step 1	4 times	
Step 5	Denature	94 °C	1.00
Step 6	Anneal	54 °C	1.00
Step 7	Elongate	72 °C	5.00
Step 8	Goto step 5	4 times	
Step 9	Denature	94 °C	1.00
Step 10	Anneal	54 °C	1.00
Step 11	Elongate	72 °C	7.00
Step 12	Goto step 9	4 times	
Step 13	Hold	4°C	
Step 14	End		

IV.h Conditions to make Downstream Right Homology Arm (DnRHA)

Reaction Components:

pBAC14567 DNA 100 ng.µl ⁻¹	10 µl
Oligo D6627 x pmol.µl ⁻¹	2.5 µl
Oligo D6626 x pmol.µl ⁻¹	2.5 µl
dNTPs 2mM	20 µl
Pfu Buffer	10 µl
dH ₂ O	54 µl
Mix and then centrifuge before adding Taq	
Taq polymerase	1 µl

Cycling parameters:

Step 1	Denature	94 °C	1.00
Step 2	Anneal	50 °C	1.00
Step 3	Elongate	72 °C	3.00
Step 4	Goto step 1	4 times	
Step 5	Denature	94 °C	1.00
Step 6	Anneal	50 °C	1.00
Step 7	Elongate	72 °C	5.00
Step 8	Goto step 5	4 times	
Step 9	Denature	94 °C	1.00
Step 10	Anneal	50 °C	1.00
Step 11	Elongate	72 °C	7.00
Step 12	Goto step 9	4 times	
Step 13	Hold	4°C	
Step 14	End		

Appendix V

Bibliography

Abremski, K., and Hoess, R. (1984). Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J Biol Chem* 259, 1509-14.

Akiyama, K., Watanabe, H., Tsukada, S., and Sasai, H. (2000). A novel method for constructing gene-targeting vectors. *Nucleic Acids Res* 28, E77.

Anguita, E., Sharpe, J. A., Sloane-Stanley, J. A., Tufarelli, C., Higgs, D. R., and Wood, W. G. (2002). Deletion of the mouse alpha-globin regulatory element (HS - 26) has an unexpectedly mild phenotype. *Blood* 100, 3450-6.

Antoch, M. P., Song, E. J., Chang, A. M., Vitaterna, M. H., Zhao, Y., Wilsbacher, L. D., Sangoram, A. M., King, D. P., Pinto, L. H., and Takahashi, J. S. (1997). Functional identification of the mouse circadian Clock gene by transgenic BAC rescue. *Cell* 89, 655-67.

Baird, J. W., Ryan, K. M., Hayes, I., Hampson, L., Heyworth, C. M., Clark, A., Wootton, M., Ansell, J. D., Menzel, U., Hole, N., and Graham, G. J. (2001). Differentiating embryonal stem cells are a rich source of haemopoietic gene products and suggest erythroid preconditioning of primitive haemopoietic stem cells. *J Biol Chem* 276, 9189-98.

Baker, M. D. (1989). High-frequency homologous recombination between duplicate chromosomal immunoglobulin mu heavy-chain constant regions. *Mol Cell Biol* 9, 5500-7.

Barinaga, M. (1994). Knockout mice: round two. *Science* 265, 26-8.

Bernet, A., Sabatier, S., Picketts, D. J., Ouazana, R., Morle, F., Higgs, D. R., and Godet, J. (1995). Targeted inactivation of the major positive regulatory element (HS-40) of the human alpha-globin gene locus. *Blood* 86, 1202-11.

Bethke, B., and Sauer, B. (1997). Segmental genomic replacement by Cre-mediated recombination: genotoxic stress activation of the p53 promoter in single-copy transformants. *Nucleic Acids Res* 25, 2828-34.

Bode, J., Schlake, T., Iber, M., Schubeler, D., Seibler, J., Snezhkov, E., and Nikolaev, L. (2000). The transgeneticist's toolbox: novel methods for the targeted modification of eukaryotic genomes. *Biol Chem* 381, 801-13.

Boris-Lawrie, K. A., and Temin, H. M. (1993). Recent advances in retrovirus vector technology. *Curr Opin Genet Dev* 3, 102-9.

Bouhassira, E. E., Kielman, M. F., Gilman, J., Fabry, M. F., Suzuka, S., Leone, O., Gikas, E., Bernini, L. F., and Nagel, R. L. (1997). Properties of the mouse alpha-globin HS-26: relationship to HS-40, the major enhancer of human alpha-globin gene expression. *Am J Hematol* 54, 30-9.

Bouhassira, E. E., Kielman, M. F., Gilman, J., Fabry, M. F., Suzuka, S., Leone, O., Gikas, E., Bernini, L. F., and Nagel, R. L. (1997). Properties of the mouse alpha-globin HS-26: relationship to HS-40, the major enhancer of human alpha-globin gene expression. *Am J Hematol* 54, 30-9.

Bouhassira, E. E., Westerman, K., and Leboulch, P. (1997). Transcriptional behavior of LCR enhancer elements integrated at the same chromosomal locus by recombinase-mediated cassette exchange. *Blood* 90, 3332-44.

Bradley, A., Evans, M., Kaufman, M. H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* 309, 255-6.

Brault, V., Moore, R., Kutsch, S., Ishibashi, M., Rowitch, D. H., McMahon, A. P., Sommer, L., Boussadia, O., and Kemler, R. (2001). Inactivation of the beta-catenin

gene by Wnt1-Cre-mediated deletion results in dramatic brain malformation and failure of craniofacial development. *Development* *128*, 1253-64.

Bucan, M., and Abel, T. (2002). The mouse: genetics meets behaviour. *Nat Rev Genet* *3*, 114-23.

Buchholz, F., Angrand, P. O., and Stewart, A. F. (1998). Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat Biotechnol* *16*, 657-62.

Buchholz, F., Refaeli, Y., Trumpp, A., and Bishop, J. M. (2000). Inducible chromosomal translocation of AML1 and ETO genes through Cre/loxP-mediated recombination in the mouse. *EMBO Rep* *1*, 133-9.

Buchholz, F., and Stewart, A. F. (2001). Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nat Biotechnol* *19*, 1047-52.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J Mol Biol* *268*, 78-94.

Campbell, K. H., McWhir, J., Ritchie, W. A., and Wilmut, I. (1996). Sheep cloned by nuclear transfer from a cultured cell line. *Nature* *380*, 64-6.

Cecconi, F., and Meyer, B. I. (2000). Gene trap: a way to identify novel genes and unravel their biological function. *FEBS Lett* *480*, 63-71.

Chang, B. H., Liao, W., Li, L., Nakamuta, M., Mack, D., and Chan, L. (1999). Liver-specific inactivation of the abetalipoproteinemia gene completely abrogates very low density lipoprotein/low density lipoprotein production in a viable conditional knockout mouse. *J Biol Chem* *274*, 6051-5.

Chartier, F. L., Keer, J. T., Sutcliffe, M. J., Henriques, D. A., Mileham, P., and Brown, S. D. (1992). Construction of a mouse yeast artificial chromosome library in a recombination-deficient strain of yeast. *Nat Genet* *1*, 132-6.

Chen, L., Kosslak, R., and Atherly, A. G. (1994). Mechanical shear of high molecular weight DNA in agarose plugs. *Biotechniques* 16, 228-9.

Choi, T. K., Hollenbach, P. W., Pearson, B. E., Ueda, R. M., Weddell, G. N., Kurahara, C. G., Woodhouse, C. S., Kay, R. M., and Loring, J. F. (1993). Transgenic mice containing a human heavy chain immunoglobulin gene fragment cloned in a yeast artificial chromosome. *Nat Genet* 4, 117-23.

Clark, A. J., and Sandler, S. J. (1994). Homologous genetic recombination: the pieces begin to fall into place. *Crit Rev Microbiol* 20, 125-42.

Clark, A. J., Sandler, S. J., Willis, D. K., Chu, C. C., Blamar, M. A., and Lovett, S. T. (1984). Genes of the RecE and RecF pathways of conjugational recombination in *Escherichia coli*. *Cold Spring Harb Symp Quant Biol* 49, 453-62.

Cleary, M. A., van Raamsdonk, C. D., Levorse, J., Zheng, B., Bradley, A., and Tilghman, S. M. (2001). Disruption of an imprinted gene cluster by a targeted chromosomal translocation in mice. *Nat Genet* 29, 78-82.

Clegg, J. B. (1987). Can the product of the theta gene be a real globin? *Nature* 329, 465-6.

Cline, J., Braman, J. C., and Hogrefe, H. H. (1996). PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res* 24, 3546-51.

Cocchia M, K. N., Kim SJ, Larionov V, Schlessinger D, Nagaraja R. (2000). Recovery and potential utility of YACs as circular YACs/BACs. *Nucleic Acids Res* 28, E81.

Colleaux, L., D'Auriol, L., Galibert, F., and Dujon, B. (1988). Recognition and cleavage site of the intron-encoded omega transposase. *Proc Natl Acad Sci U S A* 85, 6022-6.

Colman, A. (1999). Dolly, Polly and other 'ollys': likely impact of cloning technology on biomedical uses of livestock. *Genet Anal* 15, 167-73.

Costantini, F., and Lacy, E. (1981). Introduction of a rabbit beta-globin gene into the mouse germ line. *Nature* 294, 92-4.

Davies, N. P., Rosewell, I. R., and Bruggemann, M. (1992). Targeted alterations in yeast artificial chromosomes for inter-species gene transfer. *Nucleic Acids Res* 20, 2693-8.

Davies, N. P., Rosewell, I. R., Richardson, J. C., Cook, G. P., Neuberger, M. S., Brownstein, B. H., Norris, M. L., and Bruggemann, M. (1993). Creation of mice expressing human antibody light chains by introduction of a yeast artificial chromosome containing the core region of the human immunoglobulin kappa locus. *Biotechnology (N Y)* 11, 911-4.

Deng C, C. M. (1992). Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol Cell Biol* 12, 3365-3371.

Detloff, P. J., Lewis, J., John, S. W., Shehee, W. R., Langenbach, R., Maeda, N., and Smithies, O. (1994). Deletion and replacement of the mouse adult beta-globin genes by a "plug and socket" repeated targeting strategy. *Mol Cell Biol* 14, 6936-43.

Dickerson, R. E., and Geis, I. (1983). Hemoglobin : structure, function, evolution, and pathology (Menlo Park: Benjamin/Cummings Pub. Co.).

Doetschman, T., Gregg, R. G., Maeda, N., Hooper, M. L., Melton, D. W., Thompson, S., and Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* 330, 576-8.

Doetschman, T., Maeda, N., and Smithies, O. (1988). Targeted mutation of the Hprt gene in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* 85, 8583-7.

Duff, K., and Huxley, C. (1996). Targeting mutations to YACs by homologous recombination. *Methods Mol Biol* 54, 187-98.

Dunford, R., Vilageliu, L., and Moore, G. (1993). Stabilisation of a yeast artificial chromosome containing plant DNA using a recombination-deficient host. *Plant Mol Biol* 21, 1187-9.

Esperet, C., Sabatier, S., Deville, M. A., Ouazana, R., Bouhassira, E. E., Godet, J., Morle, F., and Bernet, A. (2000). Non-erythroid genes inserted on either side of human HS-40 impair the activation of its natural alpha -globin gene targets without being themselves preferentially activated. *J Biol Chem* 275, 25831-9.

Evans, M. J., and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, 154-6.

Feng, Y. Q., Seibler, J., Alami, R., Eisen, A., Westerman, K. A., Leboulch, P., Fiering, S., and Bouhassira, E. E. (1999). Site-specific chromosomal integration in mammalian cells: highly efficient CRE recombinase-mediated cassette exchange. *J Mol Biol* 292, 779-85.

Flint, J., Tufarelli, C., Peden, J., Clark, K., Daniels, R. J., Hardison, R., Miller, W., Philipsen, S., Tan-Un, K. C., McMorro, T., Frampton, J., Alter, B. P., Frischauf, A. M., and Higgs, D. R. (2001). Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster. *Hum Mol Genet* 10, 371-82.

Forget, B. G. (1993). YAC transgenes: bigger is probably better. *Proc Natl Acad Sci U S A* 90, 7909-11.

Fraser, P., Hurst, J., Collis, P., and Grosveld, F. (1990). DNaseI hypersensitive sites 1, 2 and 3 of the human beta-globin dominant control region direct position-independent expression. *Nucleic Acids Res* 18, 3503-8.

Gates, C. M., Stemmer, W. P., Kaptein, R., and Schatz, P. J. (1996). Affinity selective isolation of ligands from peptide libraries through display on a lac repressor "headpiece dimer". *J Mol Biol* 255, 373-86.

Gnirke, A., Huxley, C., Peterson, K., and Olson, M. V. (1993). Microinjection of intact 200- to 500-kb fragments of YAC DNA into mammalian cells. *Genomics* 15, 659-67.

Gordon, J. W., and Ruddle, F. H. (1981). Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science* 214, 1244-6.

Gossler, A., Doetschman, T., Korn, R., Serfling, E., and Kemler, R. (1986). Transgenesis by means of blastocyst-derived embryonic stem cell lines. *Proc Natl Acad Sci U S A* 83, 9065-9.

Grant, S. G., Jessee, J., Bloom, F. R., and Hanahan, D. (1990). Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc Natl Acad Sci U S A* 87, 4645-9.

Green, E. D., Riethman, H. C., Dutchik, J. E., and Olson, M. V. (1991). Detection and characterization of chimeric yeast artificial-chromosome clones. *Genomics* 11, 658-69.

Grosveld, F., van Assendelft, G. B., Greaves, D. R., and Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* 51, 975-85.

Gu, H., Marth, J. D., Orban, P. C., Mossmann, H., and Rajewsky, K. (1994). Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science* 265, 103-6.

Guy, L. G., Kothary, R., DeRepentigny, Y., Delvoye, N., Ellis, J., and Wall, L. (1996). The beta-globin locus control region enhances transcription of but does not confer position-independent expression onto the lacZ gene in transgenic mice. *Embo J* 15, 3713-21.

Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* 177, 4121-30.

Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hammerschmidt, M., Kane, D. A., Odenthal, J., van Eeden, F. J., Jiang, Y. J., Heisenberg, C. P., Kelsh, R. N., Furutani-Seiki, M., Vogelsang, E., Beuchle, D., Schach, U., Fabian, C., and Nusslein-Volhard, C. (1996). The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* 123, 1-36.

Haldi, M., Perrot, V., Saumier, M., Desai, T., Cohen, D., Cherif, D., Ward, D., and Lander, E. S. (1994). Large human YACs constructed in a rad52 strain show a reduced rate of chimerism. *Genomics* 24, 478-84.

Hall, S. D., Kane, M. F., and Kolodner, R. D. (1993). Identification and characterization of the *Escherichia coli* RecT protein, a protein encoded by the recE region that promotes renaturation of homologous single-stranded DNA. *J Bacteriol* 175, 277-87.

Hanazono, Y., Terao, K., Ozawa, K. (2001). Gene Transfer into Nonhuman Primate Hematopoietic Stem Cells: Implications for Gene Therapy. *Stem Cells* 19, 12-23.

Harris, N. L. (1997). Genotator: a workbench for sequence annotation. *Genome Res* 7, 754-62.

Hatton, C. S., Wilkie, A. O., Drysdale, H. C., Wood, W. G., Vickers, M. A., Sharpe, J., Ayyub, H., Pretorius, I. M., Buckle, V. J., and Higgs, D. R. (1990). Alpha-thalassemia caused by a large (62 kb) deletion upstream of the human alpha globin gene cluster. *Blood* 76, 221-7.

Higgs, D. R., Sharpe, J. A., and Wood, W. G. (1998). Understanding alpha globin gene expression: a step towards effective gene therapy. *Semin Hematol.* 35, 93-104.

Higgs, D. R., Vickers, M. A., Wilkie, A. O., Pretorius, I. M., Jarman, A. P., and Weatherall, D. J. (1989). A review of the molecular genetics of the human alpha-globin gene cluster. *Blood* 73, 1081-104.

Higgs, D. R., Wood, W. G., Jarman, A. P., Sharpe, J., Lida, J., Pretorius, I. M., and Ayyub, H. (1990). A major positive regulatory region located far upstream of the human alpha-globin gene locus. *Genes Dev* 4, 1588-601.

Higgs, D. R., Wood, W. G., Jarman, A. P., Vickers, M. A., Wilkie, A. O., Lamb, J., Vyas, P., and Bennett, J. P. (1990). The alpha-thalassemias. *Ann N Y Acad Sci* 612, 15-22.

Hill, F., Benes, V., Thomasova, D., Stewart, A. F., Kafatos, F. C., and Ansorge, W. (2000). BAC trimming: minimizing clone overlaps. *Genomics* 64, 111-3.

Hoess, R. H., Wierzbicki, A., and Abremski, K. (1986). The role of the loxP spacer region in P1 site-specific recombination. *Nucleic Acids Res* 14, 2287-300.

Hoess, R. H., Ziese, M., and Sternberg, N. (1982). P1 site-specific recombination: nucleotide sequence of the recombining sites. *Proc Natl Acad Sci U S A* 79, 3398-402.

Hooper, M., Hardy, K., Handyside, A., Hunter, S., and Monk, M. (1987). HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* 326, 292-5.

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Hammond, M., Huminiecki, L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., Mongin, E., Pettett, R., Pocock, M., Potter, S., Rust, A., Schmidt, E., Searle, S., Slater, G., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Stupka, E., Ureta-Vidal, A., Vastrik, I., and Clamp, M. (2002). The Ensembl genome database project. *Nucleic Acids Res* 30, 38-41.

Hubert, R. S., Mitchell, S., Chen, X. N., Ekmekji, K., Gadomski, C., Sun, Z., Noya, D., Kim, U. J., Chen, C., Shizuya, H., Simon, M., de Jong, P. J., and Korenberg, J. R. (1997). BAC and PAC contigs covering 3.5 Mb of the Down syndrome congenital heart disease region between D21S55 and MX1 on chromosome 21. *Genomics* 41, 218-26.

Imam, A. M., Patrinos, G. P., de Krom, M., Bottardi, S., Janssens, R. J., Katsantoni, E., Wai, A. W., Sherratt, D. J., and Grosveld, F. G. (2000). Modification of human beta-globin locus PAC clones by homologous recombination in *Escherichia coli*. *Nucleic Acids Res* 28, E65.

Izsvak, Z., Ivics, Z., and Plasterk, R. H. (2000). Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. *J Mol Biol* 302, 93-102.

Jacob, H. J., and Kwitek, A. E. (2002). Rat genetics: attaching physiology and pharmacology to the genome. *Nat Rev Genet* 3, 33-42.

Jakobovits, A., Moore, A. L., Green, L. L., Vergara, G. J., Maynard-Currie, C. E., Austin, H. A., and Klapholz, S. (1993). Germ-line transmission and expression of a human-derived yeast artificial chromosome. *Nature* 362, 255-8.

Janne, J., Hyttinen, J. M., Peura, T., Tolvanen, M., Alhonen, L., Sinervirta, R. and Halmekyto, M. (1994). Transgenic Bioreactors. *Int. J. Biochem.* 26, 859-70.

Janson, C. G., and Durning, M. J. (2001) Viral vectors as part of an integrated functional genomics program. *Genomics* 78, 3-6.

Jarman, A. P., Wood, W. G., Sharpe, J. A., Gourdon, G., Ayyub, H., and Higgs, D. R. (1991). Characterization of the major regulatory element upstream of the human alpha-globin gene cluster. *Mol Cell Biol* 11, 4679-89.

Johnson, A., Minoshima, S., Asakawa, S., Shimizu, N., Shizuya, H., Roe, B. A., and McDermid, H. E. (1999). A 1.5-Mb contig within the cat eye syndrome critical region at human chromosome 22q11.2. *Genomics* 57, 306-9.

Jones, E. A., Tosh, D., Wilson, D. I., Lindsay, S., and Forrester, L. M. (2002). Hepatic differentiation of murine embryonic stem cells. *Exp Cell Res* 272, 15-22.

Jonkers, J., and Berns, A. (2002). Conditional mouse models of sporadic cancer. *Nature Rev Cancer* 2, 251-65.

Jorgensen, E. M., and Mango, S. E. (2002). The art and design of genetic screens: *caenorhabditis elegans*. *Nat Rev Genet* 3, 356-69.

Kellendonk, C., Tronche, F., Monaghan, A. P., Angrand, P. O., Stewart, F., and Schutz, G. (1996). Regulation of Cre recombinase activity by the synthetic steroid RU 486. *Nucleic Acids Res* 24, 1404-11.

Kielman, M. F., Barradeau, S., Smits, R., Harteveld, C. L., and Bernini, L. F. (1996). Characterization and localization of the mProx1 gene directly upstream of the mouse alpha-globin gene cluster: identification of a polymorphic direct repeat in the 5'UTR. *Mamm Genome* 7, 877-80.

Kielman, M. F., Smits, R., Devi, T. S., Fodde, R., and Bernini, L. F. (1993). Homology of a 130-kb region enclosing the alpha-globin gene cluster, the alpha-locus controlling region, and two non-globin genes in human and mouse. *Mamm Genome* 4, 314-23.

Kielman, M. F., Smits, R., and Bernini, L. F. (1994). Localization and characterization of the mouse alpha-globin locus control region. *Genomics* 21, 431-3.

Kielman, M. F., Smits, R., and Bernini, L. F. (1995). Structure of the mouse 3-methyladenine DNA glycosylase gene and exact localization upstream of the alpha-globin gene cluster on chromosome 11. *Mamm Genome* 6, 499-504.

Kielman, M. F., Smits, R., Hof, I., and Bernini, L. F. (1996). Characterization and comparison of the human and mouse Dist1/alpha-globin complex reveals a tightly packed multiple gene cluster containing differentially expressed transcription units. *Genomics* 32, 341-51.

Kim, U. J., Shizuya, H., Kang, H. L., Choi, S. S., Garrett, C. L., Smink, L. J., Birren, B. W., Korenberg, J. R., Dunham, I., and Simon, M. I. (1996). A bacterial artificial chromosome-based framework contig map of human chromosome 22q. *Proc Natl Acad Sci U S A* 93, 6297-301.

Kohno, K., Oshiro, T., Kishine, H., Wada, M., Takeda, H., Ihara, N., Imamoto, F., Kano, Y., and Schlessinger, D. (1997). Construction and characterization of a *rad51rad52* double mutant as a host for YAC libraries. *Gene* 188, 175-81.

Kolb, A. F. (2001). Selection-marker-free modification of the murine beta-casein gene using a *lox2722* site. *Anal Biochem* 290, 260-71.

Kolb, A. F., and Siddell, S. G. (1997). Genomic targeting of a bicistronic DNA fragment by Cre-mediated site-specific recombination. *Gene* 203, 209-16.

Koller, U., Haas, O. A., Ludwig, W. D., Bartram, C. R., Harbott, J., Panzer-Grumayer, R., Hansen-Hagge, T., Ritter, J., Creutzig, U., Knapp, W., and et al. (1989). Phenotypic and genotypic heterogeneity in infant acute leukemia. II. Acute nonlymphoblastic leukemia. *Leukemia* 3, 708-14.

Kolossov, E., Fleischmann, B. K., Liu, Q., Bloch, W., Viatchenko-Karpinski, S., Manzke, O., Ji, G. J., Bohlen, H., Addicks, K., and Hescheler, J. (1998). Functional characteristics of ES cell-derived cardiac precursor cells identified by tissue-specific expression of the green fluorescent protein. *J Cell Biol* 143, 2045-56.

Kovesdi, I., Brough, D. E., Bruder, J. T., and Wickham, T. J. (1997). Adenoviral vectors for gene transfer. *Curr Opin Biotechnol* 8, 583-9.

Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D., and Rehrauer, W. M. (1994). Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol Rev* 58, 401-65.

Lakshmanan, G., Lieu, K. H., Grosveld, F., and Engel, J. D. (1998). Partial rescue of GATA-3 by yeast artificial chromosome transgenes. *Dev Biol* 204, 451-63.

Laloti, M., and Heath, J. (2001). A new method for generating point mutations in bacterial artificial chromosomes by homologous recombination in *Escherichia coli*. *Nucleic Acids Res* 29, E14.

Lamb, B. T., and Gearhart, J. D. (1995). YAC transgenics and the study of genetics and human disease. *Curr Opin Genet Dev* 5, 342-8.

Lamb, B. T., Sisodia, S. S., Lawler, A. M., Slunt, H. H., Kitt, C. A., Kearns, W. G., Pearson, P. L., Price, D. L., and Gearhart, J. D. (1993). Introduction and expression of the 400 kilobase amyloid precursor protein gene in transgenic mice [corrected]. *Nat Genet* 5, 22-30.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissole, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.

Larrick, J. W., and Thomas, D. W. (2001). Producing proteins in transgenic plants and animals. *Curr Opin Biotechnol* 12, 411-8.

Lee, E. C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D. A., Court, D. L., Jenkins, N. A., and Copeland, N. G. (2001). A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* 73, 56-65.

Li, X. G., Liu, D. P., and Liang, C. C. (2001). Beyond the locus control region: new light on beta-globin locus regulation. *Int J Biochem Cell Biol* 33, 914-23.

Libert, F., Lefort, A., Okimoto, R., Womack, J., and Georges, M. (1993). Construction of a bovine genomic library of large yeast artificial chromosome clones. *Genomics* 18, 270-6.

Liebhaber, S. A., Griesse, E. U., Weiss, I., Cash, F. E., Ayyub, H., Higgs, D. R., and Horst, J. (1990). Inactivation of human alpha-globin gene expression by a de novo deletion located upstream of the alpha-globin gene cluster. *Proc Natl Acad Sci U S A* 87, 9431-5.

Lindenbaum, M. H., and Grosveld, F. (1990). An in vitro globin gene switching model based on differentiated embryonic stem cells. *Genes Dev* 4, 2075-85.

Lipps, H. J., and Bode, J. (2001). Exploiting chromosomal and viral strategies: the design of safe and efficient non-viral gene transfer systems. *Curr Opin Mol Ther* 3, 133-41.

Logan, J. S. (2000). Prospects for xenotransplantation. *Curr Opin Immunol* 12, 563-8.

Lomvardas, S., and Thanos, D. (2002). Opening chromatin. *Mol Cell* 9, 209-11.

Lundberg, K. S., Shoemaker, D. D., Adams, M. W., Short, J. M., Sorge, J. A., and Mathur, E. J. (1991). High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene* 108, 1-6.

Macfarlane, W. M. (2000). Demystified.... Transcription. *Mol Pathol* 53, 1-7.

Mansour, S. L., Thomas, K. R., and Capecchi, M. R. (1988). Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* 336, 348-52.

Marks, J., Shaw, J. P., and Shen, C. K. (1986). Sequence organization and genomic complexity of primate theta 1 globin gene, a novel alpha-globin-like gene. *Nature* 321, 785-8.

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci USA* 78, 7634-8.

McLeod, M., Craft, S., and Broach, J. R. (1986). Identification of the crossover site during FLP-mediated recombination in the *Saccharomyces cerevisiae* plasmid 2 microns circle. *Mol Cell Biol* 6, 3357-67.

Means, G. D., Boyd, Y., Willis, C. R., and Derry, J. M. (2001). Transgenic rescue of the tattered phenotype by using a BAC encoding Ebp. *Mamm Genome* 12, 323-5.

Mejia, J. E., and Larin, Z. (2000). The assembly of large BACs by in vivo recombination. *Genomics* 70, 165-70.

Melton, D. W., Konecki, D. S., Brennand, J., and Caskey, C. T. (1984). Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. *Proc Natl Acad Sci U S A* 81, 2147-51.

Merscher, S., Funke, B., Epstein, J. A., Heyer, J., Puech, A., Lu, M. M., Xavier, R. J., Demay, M. B., Russell, R. G., Factor, S., Tokooya, K., Jore, B. S., Lopez, M., Pandita, R. K., Lia, M., Carrion, D., Xu, H., Schorle, H., Kobler, J. B., Scambler, P., Wynshaw-Boris, A., Skoultschi, A. I., Morrow, B. E., and Kucherlapati, R. (2001). TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome. *Cell* 104, 619-29.

Mlynarova, L., Libantova, J., Vrba, L., and Nap, J. (2002). The promiscuity of heterospecific lox sites increases dramatically in the presence of palindromic DNA. *Gene* 296, 129.

Monaco, A. P., and Larin, Z. (1994). YACs, BACs, PACs and MACs: artificial chromosomes as research tools. *Trends Biotechnol* 12, 280-6.

Monteilhet, C., Perrin, A., Thierry, A., Colleaux, L., and Dujon, B. (1990). Purification and characterization of the in vitro activity of I-Sce I, a novel and highly specific endonuclease encoded by a group I intron. *Nucleic Acids Res* 18, 1407-13.

Montoliu, L., Schedl, A., Kelsey, G., Zentgraf, H., Lichter, P., and Schutz, G. (1994). Germ line transmission of yeast artificial chromosomes in transgenic mice. *Reprod Fertil Dev* 6, 577-84.

Mortensen RM, C. D., Chao S, Geisterfer-Lowrance AA, Seidman JG. (1992). Production of homozygous mutant ES cells with a single targeting construct. *Mol Cell Biol* 12, 2391-2395.

Mullins LJ, P. C., Kotelevtseva N, Brooker G, Fleming S, Harris S, Mullins JJ. (2000). Granulation rescue and developmental marking of juxtaglomerular cells using "piggy-BAC" recombination of the mouse ren locus. *J Biol Chem* 275, 40378-40384.

Murphy, K. C. (1998). Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. *J Bacteriol* 180, 2063-71.

Muyrers, J. P., Zhang, Y., Benes, V., Testa, G., Ansorge, W., and Stewart, A. F. (2000). Point mutation of bacterial artificial chromosomes by ET recombination. *EMBO Rep* 1, 239-43.

Muyrers JP, Z. Y., Testa G, Stewart AF. (1999). Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res* 27, 1555-1557.

Nagy, A., Gocza, E., Diaz, E. M., Prideaux, V. R., Ivanyi, E., Markkula, M., and Rossant, J. (1990). Embryonic stem cells alone are able to support fetal development in the mouse. *Development* 110, 815-21.

Narayanan, K., Williamson, R., Zhang, Y., Stewart, A. F., and Ioannou, P. A. (1999). Efficient and precise engineering of a 200 kb beta-globin human/bacterial artificial chromosome in *E. coli* DH10B using an inducible homologous recombination system. *Gene Ther* 6, 442-7.

Nash, H. A., Bauer, C. E., and Gardner, J. F. (1987). Role of homology in site-specific recombination of bacteriophage lambda: evidence against joining of cohesive ends. *Proc Natl Acad Sci U S A* 84, 4049-53.

Nefedov, M., Williamson, R., and Ioannou, P. A. (2000). Insertion of disease-causing mutations in BACs by homologous recombination in *Escherichia coli*. *Nucleic Acids Res* 28, E79.

Neil, D. L., Villasante, A., Fisher, R. B., Vetrie, D., Cox, B., and Tyler-Smith, C. (1990). Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. *Nucleic Acids Res* 18, 1421-8.

Ng, P., and Baker, M. D. (1994). High-frequency loss of specific immunoglobulin production in hybridoma cell lines bearing a chromosomal immunoglobulin kappa gene modified by homologous recombination. *Somat Cell Mol Genet* 20, 107-19.

Nolan, P. M., Peters, J., Strivens, M., Rogers, D., Hagan, J., Spurr, N., Gray, I. C., Vizor, L., Brooker, D., Whitehill, E., Washbourne, R., Hough, T., Greenaway, S., Hewitt, M., Liu, X., McCormack, S., Pickford, K., Selley, R., Wells, C., Tymowska-Lalanne, Z., Roby, P., Glenister, P., Thornton, C., Thaung, C., Stevenson, J. A., Arkell, R., Mburu, P., Hardisty, R., Kiernan, A., Erven, A., Steel, K. P., Voegelings, S., Guenet, J. L., Nickols, C., Sadri, R., Nasse, M., Isaacs, A., Davies, K., Browne, M., Fisher, E. M., Martin, J., Rastan, S., Brown, S. D., and Hunter, J. (2000). A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat Genet* 25, 440-3.

Orford M, N. M., Vadolas J, Zaibak F, Williamson R, Ioannou PA. (2000). Engineering EGFP reporter constructs into a 200 kb human beta-globin BAC clone using GET Recombination. *Nucleic Acids Res* 28, E84.

Osoegawa, K., Tateno, M., Woon, P. Y., Frengen, E., Mammoser, A. G., Catanese, J. J., Hayashizaki, Y., and de Jong, P. J. (2000). Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res* 10, 116-28.

Palmieri, G., Miano, M. G., Casamassimi, A., Lania, A., Kohno, K., Schlessinger, D., D'Urso, M., and Featherstone, T. (1997). Construction of a pilot human YAC library in a recombination-defective yeast strain. *Gene* 188, 169-74.

Patton, E. E., and Zon, L. I. (2001). The art and design of genetic screens: zebrafish. *Nat Rev Genet* 2, 956-66.

Payne, C. M., Mullins, L. J., and Mullins, J. J. (1999). Manipulating large genomic clones via in vivo recombination in bacteria. *J Hum Hypertens* 13, 845-8.

Pearson, B. E., and Choi, T. K. (1993). Expression of the human beta-amyloid precursor protein gene from a yeast artificial chromosome in transgenic mice. *Proc Natl Acad Sci U S A* 90, 10578-82.

Peterson, K. R., Clegg, C. H., Li, Q., and Stamatoyannopoulos, G. (1997). Production of transgenic mice with yeast artificial chromosomes. *Trends Genet* 13, 61-6.

Peterson, K. R., Li, Q. L., Clegg, C. H., Furukawa, T., Navas, P. A., Norton, E. J., Kimbrough, T. G., and Stamatoyannopoulos, G. (1995). Use of yeast artificial chromosomes (YACs) in studies of mammalian development: production of beta-globin locus YAC mice carrying human globin developmental mutants. *Proc Natl Acad Sci U S A* 92, 5655-9.

Peterson, K. R., Zitnik, G., Huxley, C., Lowrey, C. H., Gnirke, A., Leppig, K. A., Papayannopoulou, T., and Stamatoyannopoulos, G. (1993). Use of yeast artificial chromosomes (YACs) for studying control of gene expression: correct regulation of the genes of a human beta-globin locus YAC following transfer to mouse erythroleukemia cell lines. *Proc Natl Acad Sci U S A* 90, 11207-11.

Piedrahita, J. A. (2000). Targeted modification of the domestic animal genome. *Theriogenology* 53, 105-16.

Podolsky, T., Fong, S. T., and Lee, B. T. (1996). Direct selection of tetracycline-sensitive *Escherichia coli* cells using nickel salts. *Plasmid* 36, 112-5.

Puech, A., Saint-Jore, B., Merscher, S., Russell, R. G., Cherif, D., Sirotkin, H., Xu, H., Factor, S., Kucherlapati, R., and Skoultschi, A. I. (2000). Normal cardiovascular development in mice deficient for 16 genes in 550 kb of the velocardiofacial/DiGeorge syndrome region. *Proc Natl Acad Sci U S A* 97, 10090-5.

Rabbitts, T. H., Appert, A., Chung, G., Collins, E. C., Drynan, L., Forster, A., Lobato, M. N., McCormack, M. P., Pannell, R., Spandidos, A., Stocks, M. R., Tanaka, T., and Tse, E. (2001). Mouse models of human chromosomal translocations and approaches to cancer therapy. *Blood Cells Mol Dis* 27, 249-59.

Ramirez-Solis R, L. P., Bradley A. (1995). Chromosome engineering in mice. *Nature* 378, 720-724.

Rathjen, J., and Rathjen, P. D. (2001). Mouse ES cells: experimental exploitation of pluripotent differentiation potential. *Curr Opin Genet Dev* 11, 587-94.

Reese, M. G., Kulp, D., Tammana, H., and Haussler, D. (2000). Gene--gene finding in *Drosophila melanogaster*. *Genome Res* 10, 529-38.

Reid, L. H., Gregg, R. G., Smithies, O., and Koller, B. H. (1990). Regulatory elements in the introns of the human HPRT gene are necessary for its expression in embryonic stem cells. *Proc Natl Acad Sci U S A* 87, 4299-303.

Robertson, E., Bradley, A., Kuehn, M., and Evans, M. (1986). Germ-line transmission of genes introduced into cultured pluripotent cells by retroviral vector. *Nature* 323, 445-8.

Rodriguez, C. I., Buchholz, F., Galloway, J., Sequerra, R., Kasper, J., Ayala, R., Stewart, A. F., and Dymecki, S. M. (2000). High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP. *Nat Genet* 25, 139-40.

Romao, L., Osorio-Almeida, L., Higgs, D. R., Lavinha, J., and Liebhaber, S. A. (1991). Alpha-thalassemia resulting from deletion of regulatory sequences far upstream of the alpha-globin structural genes. *Blood* 78, 1589-95.

Rossant, J. (2001). Stem cells from the Mammalian blastocyst. *Stem Cells* 19, 477-82.

Rudolph, U., and Mohler, H. (1999). Genetically modified animals in pharmacological research: future trends. *Eur J Pharmacol* 375, 327-37.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular cloning : a laboratory manual*, 2nd Edition (Cold Spring Harbor: Cold Spring Harbor Laboratory).

Sang, H. (1994). Transgenic chickens--methods and potential applications. *Trends Biotechnol* 12, 415-20.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74, 5463-7.

Sauer, B., and Henderson, N. (1988). Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc Natl Acad Sci U S A* 85, 5166-70.

Sauer, B., and Henderson, N. (1990). Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. *New Biol* 2, 441-9.

Schedl, A., Grimes, B., and Montolieu, L. (1996). YAC transfer by microinjection. *Methods Mol Biol* 54, 293-306.

Schedl, A., Larin, Z., Montoliu, L., Thies, E., Kelsey, G., Lehrach, H., and Schutz, G. (1993). A method for the generation of YAC transgenic mice by pronuclear microinjection. *Nucleic Acids Res* 21, 4783-7.

Schlake, T., and Bode, J. (1994). Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry* 33, 12746-51.

Schmidt, R., Putterill, J., West, J., Cnops, G., Robson, F., Coupland, G., and Dean, C. (1994). Analysis of clones carrying repeated DNA sequences in two YAC libraries of *Arabidopsis thaliana* DNA. *Plant J* 5, 735-44.

Sedivy, J. M., and Dutriaux, A. (1999). Gene targeting and somatic cell genetics--a rebirth or a coming of age? *Trends Genet* 15, 88-90.

Seibler, J., Schubeler, D., Fiering, S., Groudine, M., and Bode, J. (1998). DNA cassette exchange in ES cells mediated by FLP recombinase: an efficient strategy for repeated modification of tagged loci by marker-free constructs. *Biochemistry* 37, 6229-34.

Selfridge, J., Pow, A. M., McWhir, J., Magin, T. M., and Melton, D. W. (1992). Gene targeting using a mouse HPRT minigene/HPRT-deficient embryonic stem cell system: inactivation of the mouse ERCC-1 gene. *Somat Cell Mol Genet* 18, 325-36.

Sharpe, J. A., Chan-Thomas, P. S., Lida, J., Ayyub, H., Wood, W. G., and Higgs, D. R. (1992). Analysis of the human alpha globin upstream regulatory element (HS-40) in transgenic mice. *EMBO J* 11, 4565-72.

Sharpe, J. A., Wells, D. J., Whitelaw, E., Vyas, P., Higgs, D.R., and Wood, W. G. (1993). Analysis of the human alpha-globin gene cluster in transgenic mice. *Proc Natl Acad Sci U S A* 90, 11262-6.

Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments

of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci U S A* 89, 8794-7.

Slee, R., Grimes, B., Speed, R. M., Taggart, M., Maguire, S. M., Ross, A., McGill, N. I., Saunders, P. T., and Cooke, H. J. (1999). A human DAZ transgene confers partial rescue of the mouse *Dazl* null phenotype. *Proc Natl Acad Sci U S A* 96, 8040-5.

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* 336, 688-90.

Smith, A. G. (1991). Culture and differentiation of embryonic stem cells. *J Tiss Cult Meth* 13, 89-94.

Smith, A. J., De Sousa, M. A., Kwabi-Addo, B., Heppell-Parton, A., Impey, H., and Rabbitts, P. (1995). A site-directed chromosomal translocation induced in embryonic stem cells by Cre-loxP recombination. *Nat Genet* 9, 376-85.

Song, J., Dong, F., Lilly, J. W., Stupar, R. M., and Jiang, J. (2001). Instability of bacterial artificial chromosome (BAC) clones containing tandemly repeated DNA sequences. *Genome* 44, 463-9.

Soukharev, S., Miller, J. L., and Sauer, B. (1999). Segmental genomic replacement in embryonic stem cells by double lox targeting. *Nucleic Acids Res* 27, e21.

St Johnston, D. (2002). The art and design of genetic screens: *Drosophila melanogaster*. *Nat Rev Genet* 3, 176-88.

St-Onge, L., Furth, P. A., and Gruss, P. (1996). Temporal control of the Cre recombinase in transgenic mice by a tetracycline responsive promoter. *Nucleic Acids Res* 24, 3875-7.

Sternberg, N. (1981). Bacteriophage P1 site-specific recombination. III. Strand exchange during recombination at lox sites. *J Mol Biol* 150, 603-8.

Sternberg, N., and Hamilton, D. (1981). Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. *J Mol Biol* 150, 467-86.

Sternberg, N., Hamilton, D., and Hoess, R. (1981). Bacteriophage P1 site-specific recombination. II. Recombination between loxP and the bacterial chromosome. *J Mol Biol* 150, 487-507.

Steward, R., and Nusslein-Volhard, C. (1986). The genetics of the dorsal-Bicaudal-D region of *Drosophila melanogaster*. *Genetics* 113, 665-78.

Strauss, W. M., Dausman, J., Beard, C., Johnson, C., Lawrence, J. B., and Jaenisch, R. (1993). Germ line transmission of a yeast artificial chromosome spanning the murine alpha 1(I) collagen locus. *Science* 259, 1904-7.

Suzuki, A., and Nakano, T. (2001). Development of hematopoietic cells from embryonic stem cells. *Int J Hematol* 73, 1-5.

Swaminathan, S., Ellis, H. M., Waters, L. S., Yu, D., Lee, E. C., Court, D. L., and Sharan, S. K. (2001). Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis* 29, 14-21.

Tautz, D. (2000). Evolution of transcriptional regulation. *Curr Opin Genet Dev* 10, 575-9.

te Riele, H., Maandag, E. R., and Berns, A. (1992). Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proc Natl Acad Sci U S A* 89, 5128-32.

Thierry, A., Perrin, A., Boyer, J., Fairhead, C., Dujon, B., Frey, B., and Schmitz, G. (1991). Cleavage of yeast and bacteriophage T7 genomes at a single site using the rare cutter endonuclease I-Sce I. *Nucleic Acids Res* 19, 189-90.

Thomas, K. R., and Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* 51, 503-12.

Thompson, S., Clarke, A. R., Pow, A. M., Hooper, M. L., and Melton, D. W. (1989). Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell* 56, 313-21.

Tomita, S., Sinal, C. J., Yim, S. H., and Gonzalez, F. J. (2000). Conditional disruption of the aryl hydrocarbon receptor nuclear translocator (Arnt) gene leads to loss of target gene induction by the aryl hydrocarbon receptor and hypoxia-inducible factor 1alpha. *Mol Endocrinol* 14, 1674-81.

Tropepe, V., Hitoshi, S., Sirard, C., Mak, T. W., Rossant, J., and van der Kooy, D. (2001). Direct neural fate specification from embryonic stem cells: a primitive mammalian neural stem cell stage acquired through a default mechanism. *Neuron* 30, 65-78.

Valancius, V., and Smithies, O. (1991). Testing an "in-out" targeting procedure for making subtle genomic modifications in mouse embryonic stem cells. *Mol Cell Biol* 11, 1402-8.

van der Meer, I. M., Bovy, A. G., and Bosch, D. (2001). Plant-based raw material: improved food quality for better nutrition via plant genomics. *Curr Opin Biotechnol* 12, 488-92.

Van Deursen J, F. M., Van Rees B, Grosveld G. (1995). Cre-mediated site-specific translocation between nonhomologous mouse chromosomes. *Proc Natl Acad Sci* 92, 7376-7380.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M.,

Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., et al. (2001). The sequence of the human genome. *Science* 291, 1304-51.

Vickers, M. A., Vyas, P., Harris, P. C., Simmons, D. L., and Higgs, D. R. (1993) Structure of the human 3-methyladenine DNA glycosylase gene and localization close to the 16p telomere. *Proc Natl Acad Sci U S A* 90, 3437-41.

Vignal, A., Milan, D., SanCristobal, M., and Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol* 34, 275-305.

Vooijs, M., Jonkers, J., and Berns, A. (2001). A highly efficient ligand-regulated Cre recombinase mouse line shows that LoxP recombination is position dependent. *EMBO Rep* 2, 292-7.

Vyas, P., Vickers, M. A., Picketts, D. J., and Higgs, D. R. (1995). Conservation of position and sequence of a novel, widely expressed gene containing the major human alpha-globin regulatory element. *Genomics* 29, 679-89.

Wagner, K. U., Claudio, E., Rucker, E. B., 3rd, Riedlinger, G., Broussard, C., Schwartzberg, P. L., Siebenlist, U., and Hennighausen, L. (2000). Conditional deletion of the Bcl-x gene from erythroid cells results in hemolytic anemia and profound splenomegaly. *Development* 127, 4949-58.

Wakayama, T., Rodriguez, I., Perry, A. C., Yanagimachi, R., and Mombaerts, P. (1999). Mice cloned from embryonic stem cells. *Proc Natl Acad Sci U S A* 96, 14984-9.

Walters, M. C., Fiering, S., Bouhassira, E. E., Scalzo, D., Goeke, S., Magis, W., Garrick, D., Whitelaw, E., and Martin, D. I. (1999). The chicken beta-globin 5'HS4 boundary element blocks enhancer-mediated suppression of silencing. *Mol Cell Biol* 19, 3714-26.

Wang, Y., Krushel, L. A., and Edelman, G. M. (1996). Targeted DNA recombination in vivo using an adenovirus carrying the cre recombinase gene. *Proc Natl Acad Sci U S A* 93, 3932-6.

Wheeler, M. B., and Walters, E. M. (2001). Transgenic technology and applications in swine. *Theriogenology* 56, 1345-69.

Wilkie, A. O., Buckle, V. J., Harris, P. C., Lamb, J., Barton, N. J., Reeders, S. T., Lindenbaum, R. H., Nicholls, R. D., Barrow, M., Bethlenfalvay, N. C., and et al. (1990). Clinical features and molecular analysis of the alpha thalassemia/mental retardation syndromes. I. Cases due to deletions involving chromosome band 16p13.3. *Am J Hum Genet* 46, 1112-26.

Wines, M. E., Shi, Y., Lindor, M., and Holdener, B. C. (2000). Physical localization of the mesoderm development (mesd) functional region. *Genomics* 68, 322-9.

Wobus, A. M. (2001). Potential of embryonic stem cells. *Mol Aspects Med* 22, 149-64.

Yang, X. W., Model, P., and Heintz, N. (1997). Homologous recombination based modification in *Escherichia coli* and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat Biotechnol* 15, 859-65.

Yeh, R. F., Lim, L. P., and Burge, C. B. (2001). Computational inference of homologous gene structures in the human genome. *Genome Res* 11, 803-16.

Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., and Court, D. L. (2000). An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci U S A* 97, 5978-83.

Yu, Y., and Bradley, A. (2001). Engineering chromosomal rearrangements in mice. *Nat Rev Genet* 2, 780-90.

Zakany, J., and Duboule, D. (1996). Synpolydactyly in mice with a targeted deficiency in the HoxD complex. *Nature* 384, 69-71.

Zhang, Y., Buchholz, F., Muyrers, J. P., and Stewart, A. F. (1998). A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat Genet* 20, 123-8.

Zhang, Y., Muyrers, J. P., Testa, G., and Stewart, A. F. (2000). DNA cloning by homologous recombination in *Escherichia coli*. *Nat Biotechnol* 18, 1314-7.

Zhang Y, R. C., Ayral AM, Sablitzky F, Littlewood TD, Reth M. (1996). Inducible site-directed recombination in mouse embryonic stem cells. *Nucleic Acids Res* 24, 543-548.

Zhang, Z., and Lutz, B. (2002). Cre recombinase-mediated inversion using lox66 and lox71: method to introduce conditional point mutations into the CREB-binding protein. *Nucleic Acids Res* 30, e90.

Zheng, B., Sage, M., Cai, W. W., Thompson, D. M., Tavsanli, B. C., Cheah, Y. C., and Bradley, A. (1999). Engineering a mouse balancer chromosome. *Nat Genet* 22, 375-8.

Zheng, B., Sage, M., Sheppard, E. A., Jurecic, V., and Bradley, A. (2000). Engineering mouse chromosomes with Cre-loxP: range, efficiency, and somatic applications. *Mol Cell Biol* 20, 648-55.

Zimmer, R., and Verrinder Gibbins, A. M. (1997). Construction and characterization of a large-fragment chicken bacterial artificial chromosome library. *Genomics* 42, 217-26.